

Elements of Financial Risk Management, Second Edition

Peter F. Christoffersen

Elements of Financial Risk Management

Elements of Financial Risk Management

Second Edition

Peter F. Christoffersen

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
225 Wyman Street, Waltham, MA 02451, USA
The Boulevard, Langford Lane, Kidlington, Oxford, OX5 1GB, UK

Library of Congress Cataloging-in-Publication Data
Christoffersen, Peter F.
Elements of financial risk management / Peter Christoffersen. — 2nd ed.
p. cm.
ISBN 978-0-12-374448-7
1. Financial risk management. I. Title.
HD61.C548 2012
658.15′5—dc23
2011030909

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

For information on all Academic Press publications
visit our Web site at *www.elsevierdirect.com*

Printed in the United States
11 12 13 14 15 16   6 5 4 3 2 1

To Susan

Contents

Preface

Intended Readers

This book is intended for three types of readers with an interest in financial risk management: first, graduate and PhD students specializing in finance and economics; second, market practitioners with a quantitative undergraduate or graduate degree; third, advanced undergraduates majoring in economics, engineering, finance, or another quantitative field.

I have taught the less technical parts of the book in a fourth-year undergraduate finance elective course and an MBA elective on financial risk management. I covered the more technical material in a PhD course on options and risk management and in technical training courses on market risk designed for market practitioners.

In terms of prerequisites, ideally the reader should have taken as a minimum a course on investments including options, a course on statistics, and a course on linear algebra.

Software

A number of empirical exercises are listed at the end of each chapter. Excel spreadsheets with the data underlying the exercises can be found on the web site accompanying the book.

The web site also contains Excel files with answers to all the exercises. This way, virtually every technique discussed in the main text of the book is implemented in Excel using actual asset return data. The material on the web site is an essential part of the book.

Any suggestions regarding improvements to the book are most welcome. Please e-mail these suggestions to peter.christoffersen@rotman.utoronto.ca. Instructors who have adopted the book in their courses are welcome to e-mail me for a set of PowerPoint slides of the material in the book.

New in the Second Edition

The second edition of the book has five new chapters and much new material in existing chapters. The new chapters are as follows:

- Chapter 2 contains a comparison of static versus dynamic risk measures in light of the 2007–2009 financial crisis and the 1987 stock market crash.
- Chapter 3 provides an brief review of basic probability and statistics and gives a short introduction to time series econometrics.
- Chapter 5 is devoted to daily volatility models based on intraday data.
- Chapter 8 introduces nonnormal multivariate models including copula models.
- Chapter 12 gives a brief introduction to key ideas in the management of credit risk.

Organization of the Book

The new edition is organized into four parts:

- Part I provides various background material including empirical facts (Chapter 1), standard risk measures (Chapter 2), and basic statistical methods (Chapter 3).
- Part II develops a univariate risk model that allows for dynamic volatility (Chapter 4), incorporates intraday data (Chapter 5), and allows for nonnormal shocks to returns (Chapter 6).
- Part III gives a framework for multivariate risk modeling including dynamic correlations (Chapter 7), copulas (Chapter 8), and model simulation using Monte Carlo methods (Chapter 9).
- Part IV is devoted to option valuation (Chapter 10), option risk management (Chapter 11), credit risk management (Chapter 12), and finally backtesting and stress testing (Chapter 13).

For more information see the companion site at
http://www.elsevierdirect.com/companions/9780123744487

Acknowledgments

Many people have played an important part (knowingly or unknowingly) in the writing of this book. Without implication, I would like to acknowledge the following people for stimulating discussions on topics covered in this book:

My coauthors, in particular Kris Jacobs, but also Torben Andersen, Jeremy Berkowitz, Tim Bollerslev, Frank Diebold, Peter Doyle, Jan Ericsson, Vihang Errunza, Bruno Feunou, Eric Ghysels, Silvia Goncalves, Rusland Goyenko, Jinyong Hahn, Steve Heston, Atsushi Inoue, Roberto Mariano, Nour Meddahi, Amrita Nain, Denis Pelletier, Til Schuermann, Torsten Sloek, Norm Swanson, Anthony Tay, and Rob Wescott.

My Rotman School colleagues, especially John Hull, Raymond Kan, Tom McCurdy, Kevin Wang, and Alan White.

My Copenhagen Business School colleagues, especially Ken Bechman, Soeren Hvidkjaer, Bjarne Astrup Jensen, Kristian Miltersen, David Lando, Lasse Heje Pedersen, Peter Raahauge, Jesper Rangvid, Carsten Soerensen, and Mads Stenbo.

My CREATES colleagues including Ole Barndorff-Nielsen, Charlotte Christiansen, Bent Jesper Christensen, Kim Christensen, Tom Engsted, Niels Haldrup, Peter Hansen, Michael Jansson, Soeren Johansen, Dennis Kristensen, Asger Lunde, Morten Nielsen, Lars Stentoft, Timo Terasvirta, Valeri Voev, and Allan Timmermann.

My former McGill University colleagues, especially Francesca Carrieri, Benjamin Croitoru, Adolfo de Motta, and Sergei Sarkissian.

My former PhD students, especially Bo-Young Chang, Christian Dorion, Redouane Elkamhi, Xisong Jin, Lotfi Karoui, Karim Mimouni, Jaideep Oberoi, Chay Ornthanalai, Greg Vainberg, Aurelio Vasquez, and Yintian Wang.

I would also like to thank the following academics and practitioners whose work and ideas form the backbone of the book: Gurdip Bakshi, Bryan Campbell, Jin Duan, Rob Engle, John Galbraith, Rene Garcia, Eric Jacquier, Chris Jones, Michael Jouralev, Philippe Jorion, Ohad Kondor, Jose Lopez, Simone Manganelli, James MacKinnon, Saikat Nandi, Andrew Patton, Andrey Pavlov, Matthew Pritsker, Eric Renault, Garry Schinasi, Neil Shephard, Kevin Sheppard, Jean-Guy Simonato, and Jonathan Wright.

I have had a team of outstanding students working with me on the manuscript and on the Excel workbooks in particular. In the first edition they were Roustam Botachev, Thierry Koupaki, Stefano Mazzotta, Daniel Neata, and Denis Pelletier. In the second edition they are Kadir Babaoglu, Mathieu Fournier, Erfan Jafari, Hugues Langlois, and Xuhui Pan.

For financial support of my research in general and of this book in particular I would like to thank CBS, CIRANO, CIREQ, CREATES, FQRSC, IFM2, the Rotman School, and SSHRC.

I would also like to thank my editor at Academic Press, Scott Bentley, for his encouragement during the process of writing this book and Kathleen Paoni and Heather Tighe for keeping the production on track.

Finally, I would like to thank Susan for constant moral support, and Nicholas and Phillip for helping me keep perspective.

For more information see the companion site at
http://www.elsevierdirect.com/companions/9780123744487

1 Risk Management and Financial Returns

1 Chapter Outline

This chapter begins by listing the learning objectives of the book. We then ask why firms should be occupied with risk management in the first place. In answering this question, we discuss the apparent contradiction between standard investment theory and the emergence of risk management as a field, and we list theoretical reasons why managers should give attention to risk management. We also discuss the empirical evidence of the effectiveness and impact of current risk management practices in the corporate as well as financial sectors. Next, we list a taxonomy of the potential risks faced by a corporation, and we briefly discuss the desirability of exposure to each type of risk. After the risk taxonomy discussion, we define asset returns and then list the stylized facts of returns, which are illustrated by the S&P 500 equity index. We then introduce the Value-at-Risk concept. Finally, we present an overview of the remainder of the book.

2 Learning Objectives

The book is intended as a practical handbook for risk managers as well as a textbook for students. It suggests a relatively sophisticated approach to risk measurement and risk modeling. The idea behind the book is to document key features of risky asset returns and then construct tractable statistical models that capture these features. More specifically, the book is structured to help the reader

- Become familiar with the range of risks facing corporations and learn how to measure and manage these risks. The discussion will focus on various aspects of market risk.
- Become familiar with the salient features of speculative asset returns.
- Apply state-of-the-art risk measurement and risk management techniques, which are nevertheless tractable in realistic situations.

Elements of Financial Risk Management. DOI: 10.1016/B978-0-12-374448-7.00001-4

- Critically appraise commercially available risk management systems and contribute to the construction of tailor-made systems.
- Use derivatives in risk management.
- Understand the current academic and practitioner literature on risk management techniques.

3 Risk Management and the Firm

Before diving into the discussion of the range of risks facing a corporation and before analyzing the state-of-the art techniques available for measuring and managing these risks it is appropriate to start by asking the basic question about financial risk management.

3.1 Why Should Firms Manage Risk?

From a purely academic perspective, corporate interest in risk management seems curious. Classic portfolio theory tells us that investors can eliminate asset-specific risk by diversifying their holdings to include many different assets. As asset-specific risk can be avoided in this fashion, having exposure to it will not be rewarded in the market. Instead, investors should hold a combination of the risk-free asset and the market portfolio, where the exact combination will depend on the investor's appetite for risk. In this basic setup, firms should not waste resources on risk management, since investors do not care about the firm-specific risk.

From the celebrated Modigliani-Miller theorem, we similarly know that the value of a firm is independent of its risk structure; firms should simply maximize expected profits, regardless of the risk entailed; holders of securities can achieve risk transfers via appropriate portfolio allocations. It is clear, however, that the strict conditions required for the Modigliani-Miller theorem are routinely violated in practice. In particular, capital market imperfections, such as taxes and costs of financial distress, cause the theorem to fail and create a role for risk management. Thus, more realistic descriptions of the corporate setting give some justifications for why firms should devote careful attention to the risks facing them:

- *Bankruptcy costs.* The direct and indirect costs of bankruptcy are large and well known. If investors see future bankruptcy as a nontrivial possibility, then the real costs of a company reorganization or shutdown will reduce the current valuation of the firm. Thus, risk management can increase the value of a firm by reducing the probability of default.
- *Taxes.* Risk management can help reduce taxes by reducing the volatility of earnings. Many tax systems have built-in progressions and limits on the ability to carry forward in time the tax benefit of past losses. Thus, everything else being equal, lowering the volatility of future pretax income will lower the net present value of future tax payments and thus increase the value of the firm.

- *Capital structure and the cost of capital.* A major source of corporate default is the inability to service debt. Other things equal, the higher the debt-to-equity ratio, the riskier the firm. Risk management can therefore be seen as allowing the firm to have a higher debt-to-equity ratio, which is beneficial if debt financing is inexpensive net of taxes. Similarly, proper risk management may allow the firm to expand more aggressively through debt financing.
- *Compensation packages.* Due to their implicit investment in firm-specific human capital, managerial level and other key employees in a firm often have a large and unhedged exposure to the risk of the firm they work for. Thus, the riskier the firm, the more compensation current and potential employees will require to stay with or join the firm. Proper risk management can therefore help reduce the costs of retaining and recruiting key personnel.

3.2 Evidence on Risk Management Practices

A while ago, researchers at the Wharton School surveyed 2000 companies on their risk management practices, including derivatives uses. Of the 2000 firms surveyed, 400 responded. Not surprisingly, the survey found that companies use a range of methods and have a variety of reasons for using derivatives. It was also clear that not all risks that were managed were necessarily completely removed. About half of the respondents reported that they use derivatives as a risk-management tool. One-third of derivative users actively take positions reflecting their market views, thus they may be using derivatives to increase risk rather than reduce it.

Of course, not only derivatives are used to manage risky cash flows. Companies can also rely on good old-fashioned techniques such as the physical storage of goods (i.e., inventory holdings), cash buffers, and business diversification.

Not everyone chooses to manage risk, and risk management approaches differ from one firm to the next. This partly reflects the fact that the risk management goals differ across firms. In particular, some firms use cash-flow volatility, while others use the variation in the value of the firm as the risk management object of interest. It is also generally found that large firms tend to manage risk more actively than do small firms, which is perhaps surprising as small firms are generally viewed to be more risky. However, smaller firms may have limited access to derivatives markets and furthermore lack staff with risk management skills.

3.3 Does Risk Management Improve Firm Performance?

The overall answer to this question appears to be yes. Analysis of the risk management practices in the gold mining industry found that share prices were less sensitive to gold price movements after risk management. Similarly, in the natural gas industry, better risk management has been found to result in less variable stock prices. A study also found that risk management in a wide group of firms led to a reduced exposure to interest rate and exchange rate movements.

Although it is not surprising that risk management leads to lower variability—indeed the opposite finding would be shocking—a more important question is whether

risk management improves corporate performance. Again, the answer appears to be yes.

Researchers have found that less volatile cash flows result in lower costs of capital and more investment. It has also been found that a portfolio of firms using risk management would outperform a portfolio of firms that did not, when other aspects of the portfolio were controlled for. Similarly, a study found that firms using foreign exchange derivatives had higher market value than those who did not.

The evidence so far paints a fairly rosy picture of the benefits of current risk management practices in the corporate sector. However, evidence on the risk management systems in some of the largest US commercial banks is less cheerful. Several recent studies have found that while the risk forecasts on average tended to be overly conservative, perhaps a virtue at certain times, the realized losses far exceeded the risk forecasts. Importantly, the excessive losses tended to occur on consecutive days. Thus, looking back at the data on the a priori risk forecasts and the ex ante loss realizations, we would have been able to forecast an excessive loss tomorrow based on the observation of an excessive loss today. This serial dependence unveils a potential flaw in current financial sector risk management practices, and it motivates the development and implementation of new tools such as those presented in this book.

4 A Brief Taxonomy of Risks

We have already mentioned a number of risks facing a corporation, but so far we have not been precise regarding their definitions. Now is the time to make up for that.

Market risk is defined as the risk to a financial portfolio from movements in market prices such as equity prices, foreign exchange rates, interest rates, and commodity prices.

While financial firms take on a lot of market risk and thus reap the profits (and losses), they typically try to choose the type of risk to which they want to be exposed. An option trading desk, for example, has a lot of exposure to volatility changing, but not to the direction of the stock market. Option traders try to be delta neutral, as it is called. Their expertise is volatility and not market direction, and they only take on the risk about which they are the most knowledgeable, namely volatility risk. Thus financial firms tend to manage market risk actively. Nonfinancial firms, on the other hand, might decide that their core business risk (say chip manufacturing) is all they want exposure to and they therefore want to mitigate market risk or ideally eliminate it altogether.

Liquidity risk is defined as the particular risk from conducting transactions in markets with low liquidity as evidenced in low trading volume and large bid-ask spreads. Under such conditions, the attempt to sell assets may push prices lower, and assets may have to be sold at prices below their fundamental values or within a time frame longer than expected.

Traditionally, liquidity risk was given scant attention in risk management, but the events in the fall of 2008 sharply increased the attention devoted to liquidity risk. The housing crisis translated into a financial sector crises that rapidly became an equity

market crisis. The flight to low-risk treasury securities dried up liquidity in the markets for risky securities. The 2008–2009 crisis was exacerbated by a withdrawal of funding by banks to each other and to the corporate sector. Funding risk is often thought of as a type of liquidity risk.

Operational risk is defined as the risk of loss due to physical catastrophe, technical failure, and human error in the operation of a firm, including fraud, failure of management, and process errors.

Operational risk (or op risk) should be mitigated and ideally eliminated in any firm because the exposure to it offers very little return (the short-term cost savings of being careless, for example). Op risk is typically very difficult to hedge in asset markets, although certain specialized products such as weather derivatives and catastrophe bonds might offer somewhat of a hedge in certain situations. Op risk is instead typically managed using self-insurance or third-party insurance.

Credit risk is defined as the risk that a counterparty may become less likely to fulfill its obligation in part or in full on the agreed upon date. Thus credit risk consists not only of the risk that a counterparty completely defaults on its obligation, but also that it only pays in part or after the agreed upon date.

The nature of commercial banks traditionally has been to take on large amounts of credit risk through their loan portfolios. Today, banks spend much effort to carefully manage their credit risk exposure. Nonbank financials as well as nonfinancial corporations might instead want to completely eliminate credit risk because it is not part of their core business. However, many kinds of credit risks are not readily hedged in financial markets, and corporations often are forced to take on credit risk exposure that they would rather be without.

Business risk is defined as the risk that changes in variables of a business plan will destroy that plan's viability, including quantifiable risks such as business cycle and demand equation risk, and nonquantifiable risks such as changes in competitive behavior or technology. Business risk is sometimes simply defined as the types of risks that are an integral part of the core business of the firm and therefore simply should be taken on.

The risk taxonomy defined here is of course somewhat artificial. The lines between the different kinds of risk are often blurred. The securitization of credit risk via credit default swaps (CDS) is a prime example of a credit risk (the risk of default) becoming a market risk (the price of the CDS).

5 Asset Returns Definitions

While any of the preceding risks can be important to a corporation, this book focuses on various aspects of market risk. Since market risk is caused by movements in asset prices or equivalently asset returns, we begin by defining returns and then give an overview of the characteristics of typical asset returns. Because returns have much better statistical properties than price levels, risk modeling focuses on describing the dynamics of returns rather than prices.

We start by defining the daily simple rate of return from the closing prices of the asset:

$$r_{t+1} = (S_{t+1} - S_t)/S_t = S_{t+1}/S_t - 1$$

The daily continuously compounded or log return on an asset is instead defined as

$$R_{t+1} = \ln(S_{t+1}) - \ln(S_t)$$

where $\ln(*)$ denotes the natural logarithm. The two returns are typically fairly similar, as can be seen from

$$R_{t+1} = \ln(S_{t+1}) - \ln(S_t) = \ln(S_{t+1}/S_t) = \ln(1 + r_{t+1}) \approx r_{t+1}$$

The approximation holds because $\ln(x) \approx x - 1$ when x is close to 1.

The two definitions of return convey the same information but each definition has pros and cons. The simple rate of return definition has the advantage that the rate of return on a portfolio is the portfolio of the rates of return. Let N_i be the number of units (for example shares) held in asset i and let $V_{PF,t}$ be the value of the portfolio on day t so that

$$V_{PF,t} = \sum_{i=1}^{n} N_i S_{i,t}$$

Then the portfolio rate of return is

$$r_{PF,t+1} \equiv \frac{V_{PF,t+1} - V_{PF,t}}{V_{PF,t}} = \frac{\sum_{i=1}^{n} N_i S_{i,t+1} - \sum_{i=1}^{n} N_i S_{i,t}}{\sum_{i=1}^{n} N_i S_{i,t}} = \sum_{i=1}^{n} w_i r_{i,t+1}$$

where $w_i = N_i S_{i,t}/V_{PF,t}$ is the portfolio weight in asset i. This relationship does not hold for log returns because the log of a sum is not the sum of the logs.

Most assets have a lower bound of zero on the price. Log returns are more convenient for preserving this lower bound in the risk model because an arbitrarily large negative log return tomorrow will still imply a positive price at the end of tomorrow. When using log returns tomorrow's price is

$$S_{t+1} = \exp(R_{t+1}) S_t$$

where $\exp(\bullet)$ denotes the exponential function. Because the $\exp(\bullet)$ function is bounded below by zero we do not have to worry about imposing lower bounds on the distribution of returns when using log returns in risk modeling.

If we instead use the rate of return definition then tomorrow's closing price is

$$S_{t+1} = (1 + r_{t+1}) S_t$$

so that S_{t+1} could go negative in the risk model unless the assumed distribution of tomorrow's return, r_{t+1}, is bounded below by -1.

Another advantage of the log return definition is that we can easily calculate the compounded return at the $K-$day horizon simply as the sum of the daily returns:

$$R_{t+1:t+K} = \ln(S_{t+K}) - \ln(S_t) = \sum_{k=1}^{K} \ln(S_{t+k}) - \ln(S_{t+k-1}) = \sum_{k=1}^{K} R_{t+k}$$

This relationship is crucial when developing models for the term structure of interest rates and of option prices with different maturities. When using rates of return the compounded return across a $K-$day horizon involves the products of daily returns (rather than sums), which in turn complicates risk modeling across horizons.

This book will use the log return definition unless otherwise mentioned.

6 Stylized Facts of Asset Returns

We can now consider the following list of so-called stylized facts—or tendencies—which apply to most financial asset returns. Each of these facts will be discussed in detail in the book. The statistical concepts used will be explained further in Chapter 3. We will use daily returns on the S&P 500 from January 1, 2001, through December 31, 2010, to illustrate each of the features.

Daily returns have very little autocorrelation. We can write

$$Corr(R_{t+1}, R_{t+1-\tau}) \approx 0, \quad \text{for } \tau = 1, 2, 3, \ldots, 100$$

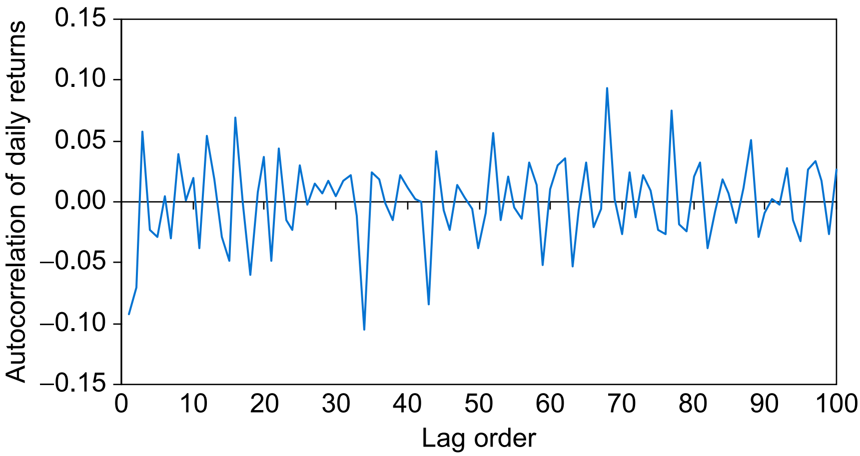
In other words, returns are almost impossible to predict from their own past. Figure 1.1 shows the correlation of daily S&P 500 returns with returns lagged from 1 to 100 days. We will take this as evidence that the conditional mean of returns is roughly constant.

The unconditional distribution of daily returns does not follow the normal distribution. Figure 1.2 shows a histogram of the daily S&P 500 return data with the normal distribution imposed. Notice how the histogram is more peaked around zero than the normal distribution. Daily returns tend to have more small positive and fewer small negative returns than the normal distribution. Although the histogram is not an ideal graphical tool for analyzing extremes, extreme returns are also more common in daily returns than in the normal distribution. We say that the daily return distribution has fat tails. Fat tails mean a higher probability of large losses (and gains) than the normal distribution would suggest. Appropriately capturing these fat tails is crucial in risk management.

The stock market exhibits occasional, very large drops but not equally large up-moves. Consequently, the return distribution is asymmetric or negatively skewed. Some markets such as that for foreign exchange tend to show less evidence of skewness.

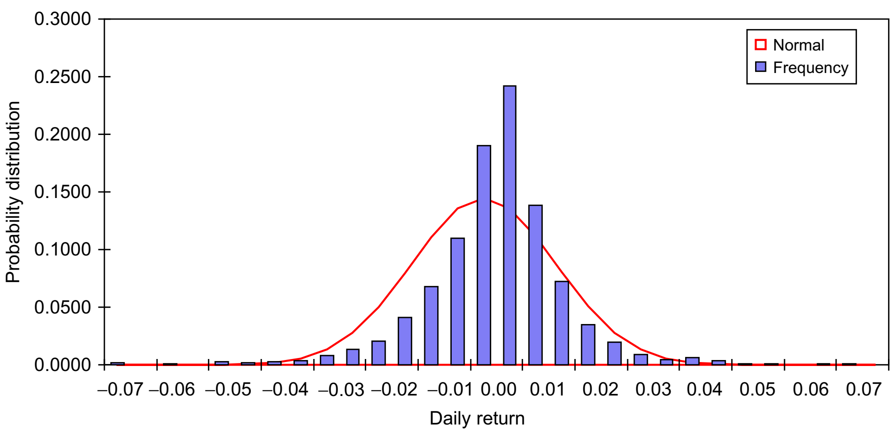
The standard deviation of returns completely dominates the mean of returns at short horizons such as daily. It is not possible to statistically reject a zero mean return. Our S&P 500 data have a daily mean of 0.0056% and a daily standard deviation of 1.3771%.

Figure 1.1 Autocorrelation of daily S&P 500 returns January 1, 2001–December 31, 2010.



Notes: Using daily returns on the S&P 500 index from January 1, 2001 through December 31, 2010, the figure shows the autocorrelations for the daily returns. The lag order on the horizontal axis refers to the number of days between the return and the lagged return for a particular autocorrelation.

Figure 1.2 Histogram of daily S&P 500 returns and the normal distribution January 1, 2001–December 31, 2010.

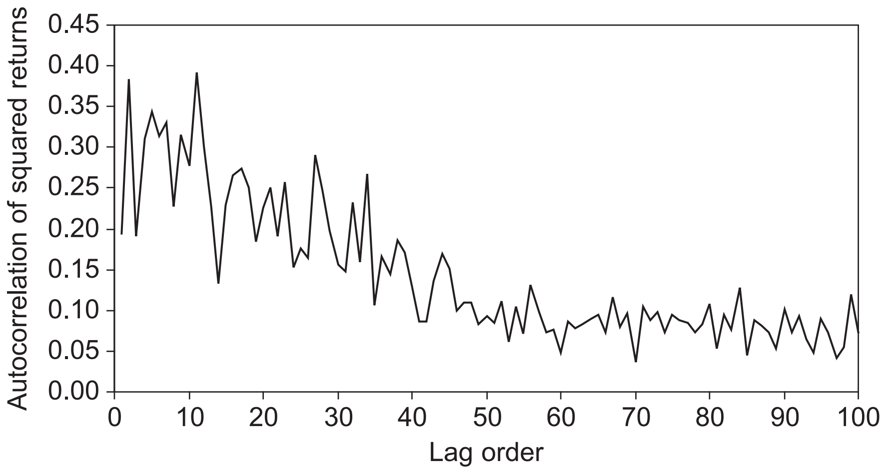


Notes: The daily S&P 500 returns from January 1, 2001 through December 31, 2010 are used to construct a histogram shown in blue bars. A normal distribution with the same mean and standard deviation as the actual returns is shown using the red line.

Variance, measured, for example, by squared returns, displays positive correlation with its own past. This is most evident at short horizons such as daily or weekly. Figure 1.3 shows the autocorrelation in squared returns for the S&P 500 data, that is

$$Corr\left(R_{t+1}^2, R_{t+1-\tau}^2\right) > 0, \quad \text{for small } \tau$$

Figure 1.3 Autocorrelation of squared daily S&P 500 returns January 1, 2010–December 31, 2010.



Notes: Using daily returns on the S&P 500 index from January 1, 2001 through December 31, 2010 the figure shows the autocorrelations for the *squared* daily returns. The lag order on the horizontal axis refers to the number of days between the squared return and the lagged squared return for a particular autocorrelation.

Models that can capture this variance dependence will be presented in Chapters 4 and 5.

Equity and equity indices display negative correlation between variance and returns. This is often called the leverage effect, arising from the fact that a drop in a stock price will increase the leverage of the firm as long as debt stays constant. This increase in leverage might explain the increase in variance associated with the price drop. We will model the leverage effect in Chapters 4 and 5.

Correlation between assets appears to be time varying. Importantly, the correlation between assets appears to increase in highly volatile down markets and extremely so during market crashes. We will model this important phenomenon in Chapter 7.

Even after standardizing returns by a time-varying volatility measure, they still have fatter than normal tails. We will refer to this as evidence of conditional nonnormality, which will be modeled in Chapters 6 and 9.

As the return-horizon increases, the unconditional return distribution changes and looks increasingly like the normal distribution. Issues related to risk management across horizons will be discussed in Chapter 8.

7 A Generic Model of Asset Returns

Based on the previous list of stylized facts, our model of individual asset returns will take the generic form

$$R_{t+1} = \mu_{t+1} + \sigma_{t+1} z_{t+1}, \quad \text{with } z_{t+1} \sim \text{i.i.d. } D(0, 1)$$

The random variable z_{t+1} is an innovation term, which we assume is identically and independently distributed (i.i.d.) according to the distribution $D(0,1)$, which has a mean equal to zero and variance equal to one. The conditional mean of the return, $E_t[R_{t+1}]$, is thus μ_{t+1}, and the conditional variance, $E_t[R_{t+1} - \mu_{t+1}]^2$, is σ^2_{t+1}.

In most of the book, we will assume that the conditional mean of the return, μ_{t+1}, is simply zero. For daily data this is a quite reasonable assumption as we mentioned in the preceding list of stylized facts. For longer horizons, the risk manager may want to estimate a model for the conditional mean as well as for the conditional variance. However, robust conditional mean relationships are not easy to find, and assuming a zero mean return may indeed be the most prudent choice the risk manager can make.

Chapters 4 and 5 will be devoted to modeling σ_{t+1}. For now we can simply rely on JP Morgan's RiskMetrics model for dynamic volatility. In that model, the volatility for tomorrow, time $t+1$, is computed at the end of today, time t, using the following simple updating rule:

$$\sigma^2_{t+1} = 0.94\sigma^2_t + 0.06R^2_t$$

On the first day of the sample, $t = 0$, the volatility σ^2_0 can be set to the sample variance of the historical data available.

8 From Asset Prices to Portfolio Returns

Consider a portfolio of n assets. The value of a portfolio at time t is again the weighted average of the asset prices using the current holdings of each asset as weights:

$$V_{PF,t} = \sum_{i=1}^{n} N_i S_{i,t}$$

The return on the portfolio between day $t+1$ and day t is then defined as

$$r_{PF,t+1} = V_{PF,t+1}/V_{PF,t} - 1$$

when using arithmetic returns, or as

$$R_{PF,t+1} = \ln\left(V_{PF,t+1}\right) - \ln\left(V_{PF,t}\right)$$

when using log returns. Note that we assume that the portfolio value on each day includes the cash from accrued dividends and other asset distributions.

Having defined the portfolio return we are ready to introduce one of the most commonly used portfolio risk measures, namely Value-at-Risk.

9 Introducing the Value-at-Risk (*VaR*) Risk Measure

Value-at-Risk, or *VaR*, is a simple risk measure that answers the following question: What loss is such that it will only be exceeded $p \cdot 100\%$ of the time in the next K

trading days? *VaR* is often defined in dollars, denoted by $\$VaR$, so that the $\$VaR$ loss is implicitly defined from the probability of getting an even larger loss as in

$$\Pr(\$Loss > \$VaR) = p$$

Note by definition that $(1-p)\,100\%$ of the time, the $\$Loss$ will be smaller than the *VaR*.

This book builds models for log returns and so we will instead use a *VaR* based on log returns defined as

$$\Pr(-R_{PF} > VaR) = p \Leftrightarrow$$
$$\Pr(R_{PF} < -VaR) = p$$

So now the $-VaR$ is defined as the number so that we would get a worse log return only with probability p. That is, we are $(1-p)\,100\%$ confident that we will get a return better than $-VaR$. This is the definition of *VaR* we will be using throughout the book. When writing the *VaR* in return terms it is much easier to gauge its magnitude. Knowing that the $\$VaR$ of a portfolio is \$500,000 does not mean much unless we know the value of the portfolio. Knowing that the return *VaR* is 15% conveys more relevant information. The appendix to this chapter shows that the two *VaR*s are related via

$$\$VaR = V_{PF}\,(1-\exp(-VaR))$$

If we start by considering a very simple example, namely that our portfolio consists of just one security, for example an S&P 500 index fund, then we can use the RiskMetrics model to provide the *VaR* for the portfolio. Let VaR^p_{t+1} denote the $p \cdot 100\%$ *VaR* for the 1-day ahead return, and assume that returns are normally distributed with zero mean and standard deviation $\sigma_{PF,t+1}$. Then

$$\begin{aligned}
\Pr\left(R_{PF,t+1} < -VaR^p_{t+1}\right) &= p \Leftrightarrow \\
\Pr\left(R_{PF,t+1}/\sigma_{PF,t+1} < -VaR^p_{t+1}/\sigma_{PF,t+1}\right) &= p \Leftrightarrow \\
\Pr\left(z_{t+1} < -VaR^p_{t+1}/\sigma_{PF,t+1}\right) &= p \Leftrightarrow \\
\Phi\left(-VaR^p_{t+1}/\sigma_{PF,t+1}\right) &= p
\end{aligned}$$

where $\Phi(*)$ denotes the cumulative density function of the standard normal distribution.

$\Phi(z)$ calculates the probability of being below the number z, and $\Phi^{-1}_p = \Phi^{-1}(p)$ instead calculates the number such that $p \cdot 100\%$ of the probability mass is below Φ^{-1}_p. Taking $\Phi^{-1}(*)$ on both sides of the preceding equation yields the *VaR* as

$$\begin{aligned}
-VaR^p_{t+1}/\sigma_{PF,t+1} &= \Phi^{-1}(p) \Leftrightarrow \\
VaR^p_{t+1} &= -\sigma_{PF,t+1}\Phi^{-1}_p
\end{aligned}$$

If we let $p = 0.01$ then we get $\Phi_p^{-1} = \Phi_{.01}^{-1} \approx -2.33$. If we assume the standard deviation forecast, $\sigma_{PF,t+1}$, for tomorrow's return is 2.5% then we get

$$\begin{aligned} VaR_{t+1}^p &= -\sigma_{PF,t+1}\Phi_p^{-1} \\ &= -0.025(-2.33) \\ &= 0.05825 \end{aligned}$$

Because Φ_p^{-1} is always negative for $p < 0.5$, the negative sign in front of the *VaR* formula again ensures that the *VaR* itself is a positive number. The interpretation is thus that the *VaR* gives a number such that there is a 1% chance of *losing* more than 5.825% of the portfolio value today. If the value of the portfolio today is \$2 million, the \$*VaR* would simply be

$$\begin{aligned} \$VaR &= V_{PF}\left(1 - \exp(-VaR)\right) \\ &= 2{,}000{,}000\left(1 - \exp(-0.05825)\right) \\ &= \$113{,}172 \end{aligned}$$

Figure 1.4 illustrates the *VaR* from a normal distribution. Notice that we assume that $K = 1$ and $p = 0.01$ here. The top panel shows the *VaR* in the probability distribution function, and the bottom panel shows the *VaR* in the cumulative distribution function. Because we have assumed that returns are normally distributed with a mean of zero, the *VaR* can be calculated very easily. All we need is a volatility forecast.

VaR has undoubtedly become the industry benchmark for risk calculation. This is because it captures an important aspect of risk, namely how bad things can get with a certain probability, p. Furthermore, it is easily communicated and easily understood.

VaR does, however, have drawbacks. Most important, extreme losses are ignored. The *VaR* number only tells us that 1% of the time we will get a return below the reported *VaR* number, but it says nothing about what will happen in those 1% worst cases. Furthermore, the *VaR* assumes that the portfolio is constant across the next K days, which is unrealistic in many cases when K is larger than a day or a week. Finally, it may not be clear how K and p should be chosen. Later we will discuss other risk measures that can improve on some of the shortcomings of *VaR*.

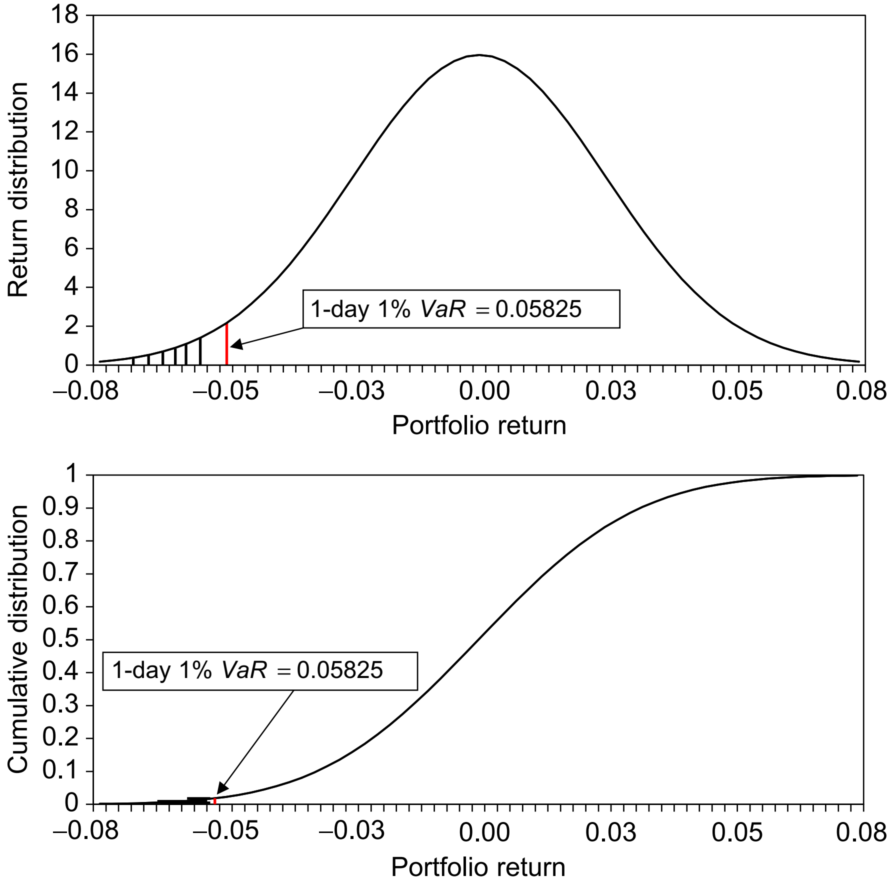
As another simple example, consider a portfolio whose value consists of 40 shares in Microsoft (MS) and 50 shares in GE. A simple way to calculate the *VaR* for the portfolio of these two stocks is to collect historical share price data for MS and GE and construct the historical portfolio pseudo returns using

$$\begin{aligned} R_{PF,t+1} &= \ln\left(V_{PF,t+1}\right) - \ln\left(V_{PF,t}\right) \\ &= \ln\left(40S_{MS,t+1} + 50S_{GE,t+1}\right) - \ln\left(40S_{MS,t} + 50S_{GE,t}\right) \end{aligned}$$

where the stock prices include accrued dividends and other distributions. Constructing a time series of past portfolio pseudo returns enables us to generate a portfolio volatility series using for example the RiskMetrics approach where

$$\sigma_{PF,t+1}^2 = 0.94\sigma_{PF,t}^2 + 0.06R_{PF,t}^2$$

Figure 1.4 Value at Risk (*VaR*) from the normal distribution return probability distribution (top panel) and cumulative return distribution (bottom panel).



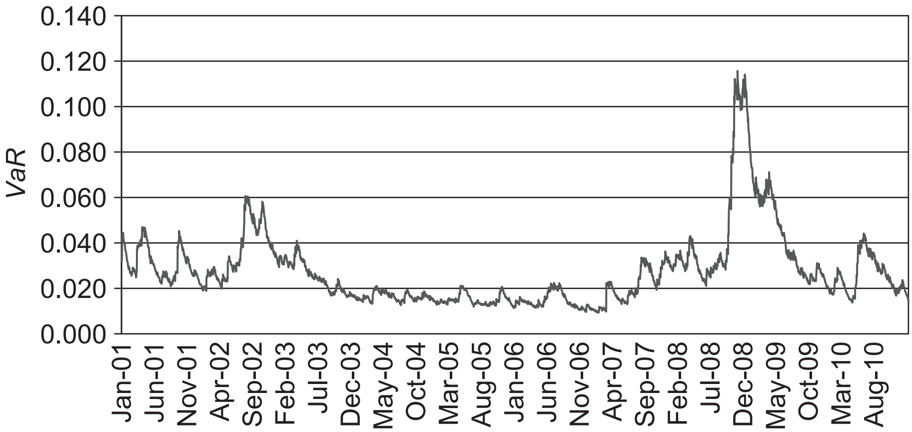
Notes: The top panel shows the probability density function of a normal distribution with a mean of zero and a standard deviation of 2.5%. The 1-day, 1% *VaR* is indicated on the horizontal axis. The bottom panel shows the cumulative density function for the same normal distribution.

We can now directly model the volatility of the portfolio return, $R_{PF,t+1}$, call it $\sigma_{PF,t+1}$, and then calculate the *VaR* for the portfolio as

$$VaR_{t+1}^{p} = -\sigma_{PF,t+1}\Phi_p^{-1}$$

where we assume that the portfolio returns are normally distributed. Figure 1.5 shows this *VaR* plotted over time. Notice that the *VaR* can be relatively low for extended periods of time but then rises sharply when volatility is high in the market, for example during the corporate defaults including the WorldCom bankruptcy in the summer of 2002 and during the financial crisis in the fall of 2008.

Figure 1.5 1-day, 1% *VaR* using RiskMetrics in S&P 500 portfolio January 1, 2001–December 31, 2010.



Notes: The daily 1-day, 1% *VaR* is plotted during the 2001–2010 period. The *VaR* is computed using a return mean of zero, using the RiskMetrics model for variance, and using a normal distribution for the return shocks.

Notice that this aggregate *VaR* method is directly dependent on the portfolio positions (40 shares and 50 shares), and it would require us to redo the volatility modeling every time the portfolio is changed or every time we contemplate change and want to study the impact on *VaR* of changing the portfolio allocations. Although modeling the aggregate portfolio return directly may be appropriate for passive portfolio risk measurement, it is not as useful for active risk management. To do sensitivity analysis and assess the benefits of diversification, we need models of the dependence between the return on individual assets or risk factors. We will consider univariate, portfolio-level risk models in Part II of the book and multivariate or asset level risk models in Part III of the book.

We also hasten to add that the assumption of normality when computing *VaR* is made for convenience and is not realistic. Important methods for dealing with the nonnormality evident in daily returns will be discussed in Chapter 6 of Part II (univariate nonnormality) and in Chapter 9 of Part III (multivariate nonnormality).

10 Overview of the Book

The book is split into four parts and contains a total of 13 chapters including this one.

Part I, which includes Chapters 1 through 3, contains various background material on risk management. Chapter 1 has discussed the motivation for risk management and listed important stylized facts that the risk model should capture. Chapter 2 introduces the Historical Simulation approach to Value-at-Risk and discusses the reasons for going beyond the Historical Simulation approach when measuring risk.

Chapter 2 also compares the Value-at-Risk and Expected Shortfall risk measures. Chapter 3 provides a primer on the basic concepts in probability and statistics used in financial risk management. It can be skipped by readers with a strong statistical background.

Part II of the book includes Chapters 4 through 6 and develops a framework for risk measurement at the portfolio level. All the models introduced in Part II are univariate. They can be used to model assets individually or to model the aggregate portfolio return. Chapter 4 discusses methods for estimating and forecasting time-varying daily return variance using daily data. Chapter 5 uses intraday return data to model and forecast daily variance. Chapter 6 introduces methods to model the tail behavior in asset returns that is not captured by volatility models and that is not captured by the normal distribution.

Part III includes Chapters 7 through 9 and it covers multivariate risk models that are capable of aggregating asset level risk models to provide sensible portfolio level risk measures. Chapter 7 introduces dynamic correlation models, which together with the dynamic volatility models in Chapters 4 and 5 can be used to construct dynamic covariance matrices for many assets. Chapter 9 introduces copula models that can be used to aggregate the univariate distribution models in Chapter 6 and thus provide proper multivariate distributions. Chapter 8 shows how the various models estimated on daily data can be used via simulation to provide estimates of risk across different investment horizons.

Part IV of the book includes Chapters 10 through 13 and contains various further topics in risk management. Chapter 10 develops models for pricing options when volatility is dynamic. Chapter 11 discusses the risk management of portfolios that include options. Chapter 12 discusses credit risk management. Chapter 13 develops methods for backtesting and stress testing risk models.

Appendix: Return *VaR* and $*VaR*

This appendix shows the relationship between the return *VaR* using log returns and the $*VaR*. First, the unknown future value of the portfolio is $V_{PF}\exp(R_{PF})$ where V_{PF} is the current market value of the portfolio and R_{PF} is log return on the portfolio. The dollar loss $*Loss* is simply the negative change in the portfolio value and so the relationship between the portfolio log return R_{PF} and the $*Loss* is

$$\$Loss = V_{PF}\left(1 - \exp\left(R_{PF}\right)\right)$$

Substituting this relationship into the definition of the $*VaR* yields

$$\Pr\left(V_{PF}\left(1 - \exp\left(R_{PF}\right)\right) > \$VaR\right) = p$$

Solving for R_{PF} yields

$$\begin{aligned}&\Pr\left(1 - \exp\left(R_{PF}\right) > \$VaR/V_{PF}\right) = p \Leftrightarrow\\&\quad \Pr\left(R_{PF} < \ln\left(1 - \$VaR/V_{PF}\right)\right) = p\end{aligned}$$

This gives us the relationship between the two *VaR*s

$$VaR = -\ln(1 - \$VaR/V_{PF})$$

or equivalently

$$\$VaR = V_{PF}(1 - \exp(-VaR))$$

Further Resources

A very nice review of the theoretical and empirical evidence on corporate risk management can be found in Stulz (1996) and Damodaran (2007).

For empirical evidence on the efficacy of risk management across a range of industries, see Allayannis and Weston (2003), Cornaggia (2010), MacKay and Moeller (2007), Minton and Schrand (1999), Purnanandam (2008), Rountree et al. (2008), Smithson (1999), and Tufano (1998).

Berkowitz and O'Brien (2002), Perignon and Smith (2010a, 2010b), and Perignon et al. (2008) document the performance of risk management systems in large commercial banks, and Dunbar (1999) contains a discussion of the increased focus on risk management after the turbulence in the fall of 1998.

The definitions of the main types of risk used here can be found at www.erisk.com and in JPMorgan/Risk Magazine (2001).

The stylized facts of asset returns are provided in Cont (2001). Surveys of Value-at-Risk models include Andersen et al. (2006), Basle Committee for Banking Supervision (2011), Christoffersen (2009), Duffie and Pan (1997), Kuester et al. (2006), and Marshall and Siegel (1997).

Useful web sites include www.gloriamundi.org, www.risk.net, www.defaultrisk.com, and www.bis.org. See also www.christoffersen.com.

References

Allayannis, G., Weston, J., 2003. Earnings Volatility, Cash-Flow Volatility and Firm Value. Manuscript, University of Virginia, Charlottesville, VA and Rice University, Houston, TX.

Andersen, T.G., Bollerslev, T., Christoffersen, P.F., Diebold, F.X., 2006. Practical Volatility and Correlation Modeling for Financial Market Risk Management. In: Carey, M., Stulz, R. (Eds.), The NBER Volume on Risks of Financial Institutions, University of Chicago Press, Chicago, IL.

Basle Committee for Banking Supervision, 2011. Messages from the Academic Literature on Risk Measurement for the Trading Book. Basel Committee on Banking Supervision, Working Paper, Basel, Switzerland.

Berkowitz, J., O'Brien, J., 2002. How accurate are Value-at-Risk models at commercial banks? J. Finance 57, 1093–1112.

Christoffersen, P.F., 2009. Value-at-Risk models. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P., Mikosch, T. (Eds.), Handbook of Financial Time Series, Springer Verlag, Düsseldorf, Germany.

Cont, R., 2001. Empirical properties of asset returns: Stylized facts and statistical issues. Quant. Finance 1, 223–236.
Cornaggia, J., 2010. Does Risk Management Matter? Manuscript, Indiana University, Bloomington, IN.
Damodaran, A., 2007. Strategic Risk Taking: A Framework for Risk Management. Wharton School Publishing, Pearson Education, Inc. Publishing as Prentice Hall, Upper Saddle River, NY.
Duffie, D., Pan, J., 1997. An overview of Value-at-Risk. J. Derivatives 4, 7–49.
Dunbar, N., 1999. The new emperors of Wall Street. Risk 26–33.
JPMorgan/Risk Magazine, 2001. Guide to Risk Management: A Glossary of Terms. Risk Waters Group, London.
Kuester, K., Mittnik, S., Paolella, M.S., 2006. Value-at-Risk prediction: A comparison of alternative strategies. J. Financial Econom. 4, 53–89.
MacKay, P., Moeller, S.B., 2007. The value of corporate risk management. J. Finance 62, 1379–1419.
Marshall, C., Siegel, M., 1997. Value-at-Risk: Implementing a risk measurement standard. J. Derivatives 4, 91–111.
Minton, B., Schrand, C., 1999. The impact of cash flow volatility on discretionary investment and the costs of debt and equity financing. J. Financial Econom. 54, 423–460.
Perignon, C., Deng, Z., Wang, Z., 2008. Do banks overstate their Value-at-Risk? J. Bank. Finance 32, 783–794.
Perignon, C., Smith, D., 2010a. The level and quality of Value-at-Risk disclosure by commercial banks. J. Bank. Finance 34, 362–377.
Perignon, C., Smith, D., 2010b. Diversification and Value-at-Risk. J. Bank. Finance 34, 55–66.
Purnanandam, A., 2008. Financial distress and corporate risk management: Theory and evidence. J. Financial Econom. 87, 706–739.
Rountree, B., Weston, J.P., Allayannis, G., 2008. Do investors value smooth performance? J. Financial Econom. 90, 237–251.
Smithson, C., 1999. Does risk management work? Risk 44–45.
Stulz, R., 1996. Rethinking risk management. J. Appl. Corp. Finance 9, 8–24.
Tufano, P., 1998. The determinants of stock price exposure: Financial engineering and the gold mining industry. J. Finance 53, 1015–1052.

Empirical Exercises

Open the Chapter1Data.xlsx file on the web site. (*Excel hint*: Enable the Data Analysis Tool under Tools, Add-Ins.)

1. From the S&P 500 prices, remove the prices that are simply repeats of the previous day's price because they indicate a missing observation due to a holiday. Calculate daily log returns as $R_{t+1} = \ln(S_{t+1}) - \ln(S_t)$ where S_{t+1} is the closing price on day $t+1$, S_t is the closing price on day t, and $\ln(*)$ is the natural logarithm. Plot the closing prices and returns over time.

2. Calculate the mean, standard deviation, skewness, and kurtosis of returns. Plot a histogram of the returns with the normal distribution imposed as well. (*Excel hints:* You can either use the Histogram tool under Data Analysis, or you can use the functions AVERAGE, STDEV, SKEW, KURT, and the array function FREQUENCY, as well as the NORMDIST function. Note that KURT computes excess kurtosis.)

3. Calculate the first through 100th lag autocorrelation. Plot the autocorrelations against the lag order. (*Excel hint*: Use the function CORREL.) Compare your result with Figure 1.1.
4. Calculate the first through 100th lag autocorrelation of squared returns. Again, plot the autocorrelations against the lag order. Compare your result with Figure 1.3.
5. Set σ_0^2 (i.e., the variance of the first observation) equal to the variance of the entire sequence of returns (you can square the standard deviation found earlier). Then calculate $\sigma_{t+1}^2 = 0.94\sigma_t^2 + 0.06R_t^2$ for $t = 2, 3, \ldots, T$ (the last observation). Plot the sequence of standard deviations (i.e., plot σ_t).
6. Compute standardized returns as $z_t = R_t/\sigma_t$ and calculate the mean, standard deviation, skewness, and kurtosis of the standardized returns. Compare them with those found in exercise 2.
7. Calculate daily, 5-day, 10-day, and 15-day nonoverlapping log returns. Calculate the mean, standard deviation, skewness, and kurtosis for all four return horizons. Do the returns look more normal as the horizon increases?
8. Calculate the 1-day, 1% *VaR* on each day in the sample using the sequence of variances σ_{t+1}^2 and the standard normal distribution assumption for the shock z_{t+1}.

The answers to these exercises can be found in the Chapter1Results.xlsx file on the companion site.

For more information see the companion site at
http://www.elsevierdirect.com/companions/9780123744487

2 Historical Simulation, Value-at-Risk, and Expected Shortfall

1 Chapter Overview

The main objectives of this chapter are twofold. First we want to introduce the most commonly used method for computing *VaR*, Historical Simulation, and we discuss the pros and cons of this method. We then discuss the pros and cons of the *VaR* risk measure itself and consider the Expected Shortfall (*ES*) alternative.

The chapter is organized as follows:

- We introduce the Historical Simulation (HS) method and discuss its pros and particularly its cons.
- We consider an extension of HS, often referred to as Weighted Historical Simulation (WHS). We compare HS and WHS during the 1987 crash.
- We then study the performance of HS and RiskMetrics during the 2008–2009 financial crisis.
- We simulate artificial return data and assess the HS *VaR* on this data.
- Finally we compare the *VaR* risk measure with a potentially more informative alternative, *ES*.

The overall conclusion from this chapter is that HS is problematic for computing *VaR*. This will motivate the dynamic models considered later. These models can be used to compute Expected Shortfall or any other desired risk measure.

2 Historical Simulation

This section defines the HS approach to Value-at-Risk and then discusses the pros and cons of the approach.

Elements of Financial Risk Management. DOI: 10.1016/B978-0-12-374448-7.00002-6

2.1 Defining Historical Simulation

Let today be day t. Consider a portfolio of n assets. If we today own $N_{i,t}$ units or shares of asset i then the value of the portfolio today is

$$V_{PF,t} = \sum_{i=1}^{n} N_{i,t} S_{i,t}$$

Using today's portfolio holdings but historical asset prices we can compute the history of "pseudo" portfolio values that would have materialized if today's portfolio allocation had been used through time. For example, yesterday's pseudo portfolio value is

$$V_{PF,t-1} = \sum_{i=1}^{n} N_{i,t} S_{i,t-1}$$

This is a pseudo value because the units of each asset held typically changes over time. The pseudo log return can now be defined as

$$R_{PF,t} = \ln\left(V_{PF,t}/V_{PF,t-1}\right)$$

Armed with this definition, we are now ready to define the Historical Simulation approach to risk management. The HS technique is deceptively simple. Consider the availability of a past sequence of m daily hypothetical portfolio returns, calculated using past prices of the underlying assets of the portfolio, but using today's portfolio weights; call it $\left\{R_{PF,t+1-\tau}\right\}_{\tau=1}^{m}$.

The HS technique simply assumes that the distribution of tomorrow's portfolio returns, $R_{PF,t+1}$, is well approximated by the empirical distribution of the past m observations, $\left\{R_{PF,t+1-\tau}\right\}_{\tau=1}^{m}$. Put differently, the distribution of $R_{PF,t+1}$ is captured by the histogram of $\left\{R_{PF,t+1-\tau}\right\}_{\tau=1}^{m}$. The *VaR* with coverage rate, p, is then simply calculated as $100p$th percentile of the sequence of past portfolio returns. We write

$$VaR_{t+1}^{p} = -Percentile\left(\left\{R_{PF,t+1-\tau}\right\}_{\tau=1}^{m}, 100p\right)$$

Thus, we simply sort the returns in $\left\{R_{PF,t+1-\tau}\right\}_{\tau=1}^{m}$ in ascending order and choose the VaR_{t+1}^{p} to be the number such that only $100p\%$ of the observations are smaller than the VaR_{t+1}^{p}. As the *VaR* typically falls in between two observations, linear interpolation can be used to calculate the exact number. Standard quantitative software packages will have the *Percentile* or similar functions built in so that the linear interpolation is performed automatically.

2.2 Pros and Cons of Historical Simulation

Historical Simulation is widely used in practice. The main reasons are (1) the ease with which is it implemented and (2) its model-free nature.

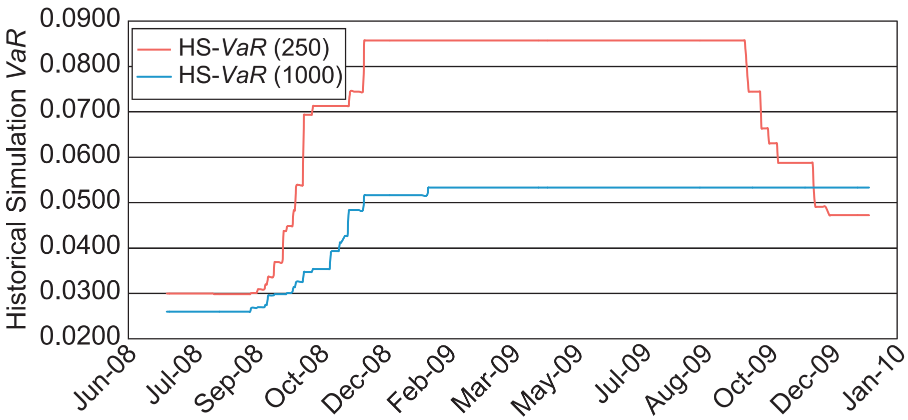
The first advantage is difficult to argue with. The HS technique clearly is very easy to implement. No parameters have to be estimated by maximum likelihood or any other method. Therefore, no numerical optimization has to be performed.

The second advantage is more contentious, however. The HS technique is model-free in the sense that it does not rely on any particular parametric model such as a RiskMetrics model for variance and a normal distribution for the standardized returns. HS lets the past m data points speak fully about the distribution of tomorrow's return without imposing any further assumptions. Model-free approaches have the obvious advantage compared with model-based approaches that relying on a model can be misleading if the model is poor.

The model-free nature of the HS model also has serious drawbacks, however.

Consider the choice of the data sample length, m. How large should m be? If m is too large, then the most recent observations, which presumably are the most relevant for tomorrow's distribution, will carry very little weight, and the *VaR* will tend to look very smooth over time. If m is chosen to be too small, then the sample may not include enough large losses to enable the risk manager to calculate, say, a 1% *VaR* with any precision. Conversely, the most recent past may be very unusual, so that tomorrow's *VaR* will be too extreme. The upshot is that the choice of m is very ad hoc, and, unfortunately, the particular choice of m matters a lot for the magnitude and dynamics of *VaR* from the HS technique. Typically m is chosen in practice to be between 250 and 1000 days corresponding to approximately 1 to 4 years. Figure 2.1 shows *VaR*s from HS $m = 250$ and $m = 1000$, respectively, using daily returns on the S&P 500 for July 1, 2008 through December 31, 2009. Notice the curious box-shaped patterns that arise from the abrupt inclusion and exclusion of large losses in the moving sample.

Figure 2.1 *VaR*s from Historical Simulation using 250 and 1,000 return days: July 1, 2008–December 31, 2009.



Notes: Daily returns on the S&P 500 index are used to compute 1-day, 1% *VaR* on a moving window of returns. The red line uses 250 days in the moving window and the blue line uses 1,000 days.

Notice also how the dynamic patterns of the HS *VaR*s are crucially dependent on *m*. The 250-day HS *VaR* is almost twice as high as the 1000-day *VaR* during the crisis period. Furthermore, the 250-day *VaR* rises quicker at the beginning of the crisis and it drops quicker as well at the end of the crisis. The key question is whether the HS *VaR* rises quickly enough and to the appropriate level.

The lack of properly specified dynamics in the HS methodology causes it to ignore well-established stylized facts on return dependence, most importantly variance clustering. This typically causes the HS *VaR* to react too slowly to changes in the market risk environment. We will consider a stark example of this next.

Because a reasonably large *m* is needed in order to calculate 1% *VaR*s with any degree of precision, the HS technique has a serious drawback when it comes to calculating the *VaR* for the next, say, 10 days rather than the next day. Ideally, the 10-day *VaR* should be calculated from 10-day nonoverlapping past returns, which would entail coming up with 10 times as many past daily returns. This is often not feasible. Thus, the model-free advantage of the HS technique is simultaneously a serious drawback. As the HS method does not rely on a well-specified dynamic model, we have no theoretically correct way of extrapolating from the 1-day distribution to get the 10-day distribution other than finding more past data. While it may be tempting to simply multiply the 1-day *VaR* from HS by $\sqrt{10}$ to obtain a 10-day *VaR*, doing so is only valid under the assumption of normality, which the HS approach is explicitly tailored to avoid.

In contrast, the dynamic return models suggested later in the book can be generalized to provide return distributions at any horizon. We will consider methods to do so in Chapter 8.

3 Weighted Historical Simulation (WHS)

We have discussed the inherent tension in the HS approach regarding the choice of sample size, *m*. If *m* is too small, then we do not have enough observations in the left tail to calculate a precise *VaR* measure, and if *m* is too large, then the *VaR* will not be sufficiently responsive to the most recent returns, which presumably have the most information about tomorrow's distribution.

We now consider a modification of the HS technique, which is designed to relieve the tension in the choice of *m* by assigning relatively more weight to the most recent observations and relatively less weight to the returns further in the past. This technique is referred to as Weighted Historical Simulation (WHS).

WHS is implemented as follows:

- Our sample of *m* past hypothetical returns, $\left\{R_{PF,t+1-\tau}\right\}_{\tau=1}^{m}$, is assigned probability weights declining exponentially through the past as follows:

$$\eta_\tau = \left\{\eta^{\tau-1}(1-\eta)/\left(1-\eta^m\right)\right\}_{\tau=1}^{m}$$

  so that, for example, today's observation is assigned the weight $\eta_1 = (1-\eta)/(1-\eta^m)$. Note that η_τ goes to zero as τ gets large, and that the weights η_τ for $\tau = 1, 2.., m$ sum to 1.

  Typically, η is assumed to be a number between 0.95 and 0.99.

- The observations along with their assigned weights are sorted in ascending order.
- The $100p\%$ *VaR* is calculated by accumulating the weights of the ascending returns until $100p\%$ is reached. Again, linear interpolation can be used to calculate the exact *VaR* number between the two sorted returns with cumulative probability weights surrounding p.

Notice that once η is chosen, the WHS technique still does not require estimation and thus retains the ease of implementation, which is the hallmark of simple HS. It has the added advantage that the weighting function builds dynamics into the technique: Today's market conditions matter more because today's return gets weighted much more than past returns. The weighting function also makes the choice of m somewhat less crucial.

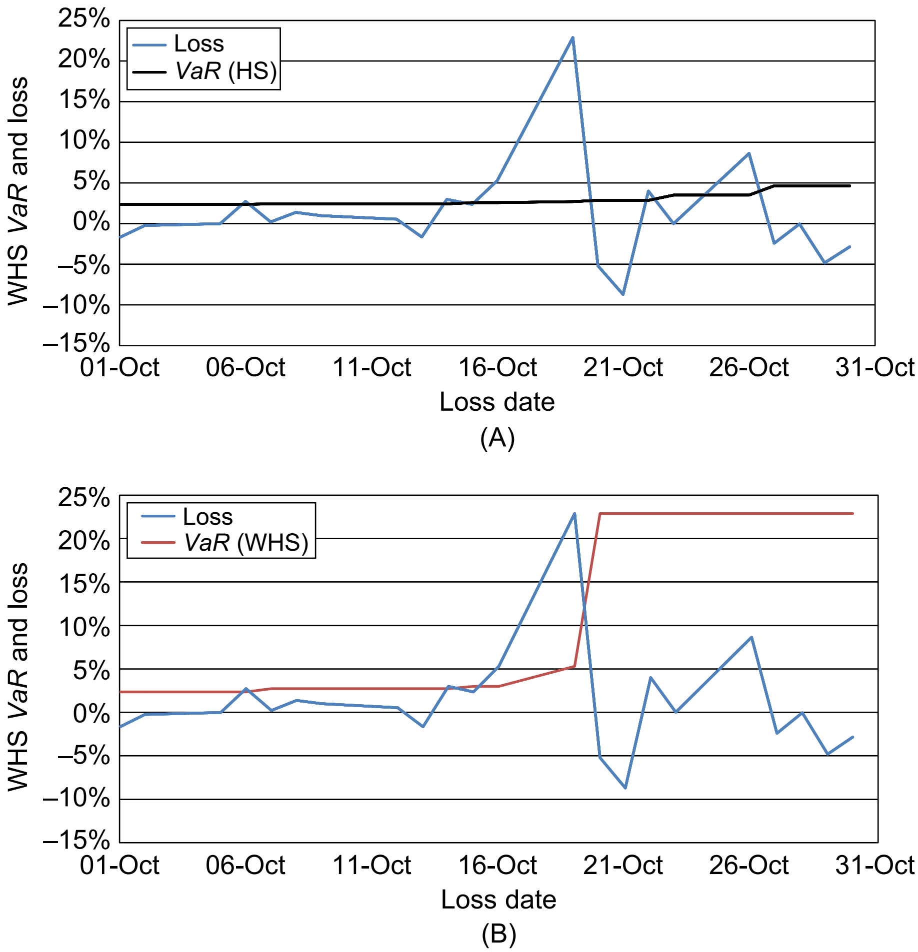
An obvious downside of the WHS approach is that no guidance is given on how to choose η. A more subtle, but also much more important downside is the effect on the weighting scheme of positive versus negative past returns—a downside that WHS shares with HS. We illustrate this with a somewhat extreme example drawing on the month surrounding the October 19, 1987, crash in the stock market. Figure 2.2 contains two panels both showing in blue lines the daily losses on a portfolio consisting of a \$1 long position in the S&P 500 index. Notice how the returns are relatively calm before October 19, when a more than 20% loss from the crash set off a dramatic increase in market variance.

The blue line in the top panel shows the *VaR* from the simple HS technique, using an m of 250. The key thing to notice of course is how the simple HS technique responds slowly and relatively little to the dramatic loss on October 19. The HS's lack of response to the crash is due to its static nature: Once the crash occurs, it simply becomes another observation in the sample that carries the same weight as the other 250 past observations. The *VaR* from the WHS method in the bottom panel (shown in red) shows a much more rapid and large response to the *VaR* forecast from the crash. As soon as the large portfolio loss from the crash is recorded, it gets assigned a large weight in the weighting scheme, which in turn increases the *VaR* dramatically. The WHS *VaR*s in Figure 2.2 assume a η of 0.99.

Thus, apparently the WHS performs its task sublimely. The dynamics of the weighting scheme kicks in to lower the *VaR* exactly when our intuition says it should. Unfortunately, all is not well. Consider Figure 2.3, which in both panels shows the daily losses from a short \$1 position in the S&P 500 index. Thus, we have simply flipped the losses from before around the x-axis. The top panel shows the *VaR* from HS, which is even more sluggish than before: Since we are short the S&P 500, the market crash corresponds to a large gain rather than a large loss. Consequently, it has no impact on the *VaR*, which is calculated from the largest losses only. Consider now the WHS *VaR* instead. The bottom panel of Figure 2.3 shows that as we are short the market, the October 19 crash has no impact on our *VaR*, only the subsequent market rebound, which corresponds to a loss for us, increases the *VaR*.

Thus, the upshot is that while WHS responds quickly to large losses, it does not respond to large gains. Arguably it should. The market crash sets off an increase in market variance, which the WHS only picks up if the crash is bad for our portfolio position. To put it bluntly, the WHS treats a large loss as a signal that risk has

Figure 2.2 (A) Historical Simulation *VaR* and daily losses from **Long** S&P 500 position, October 1987. (B) Weighted Historical Simulation *VaR* and daily losses from **Long** S&P 500 position, October 1987.

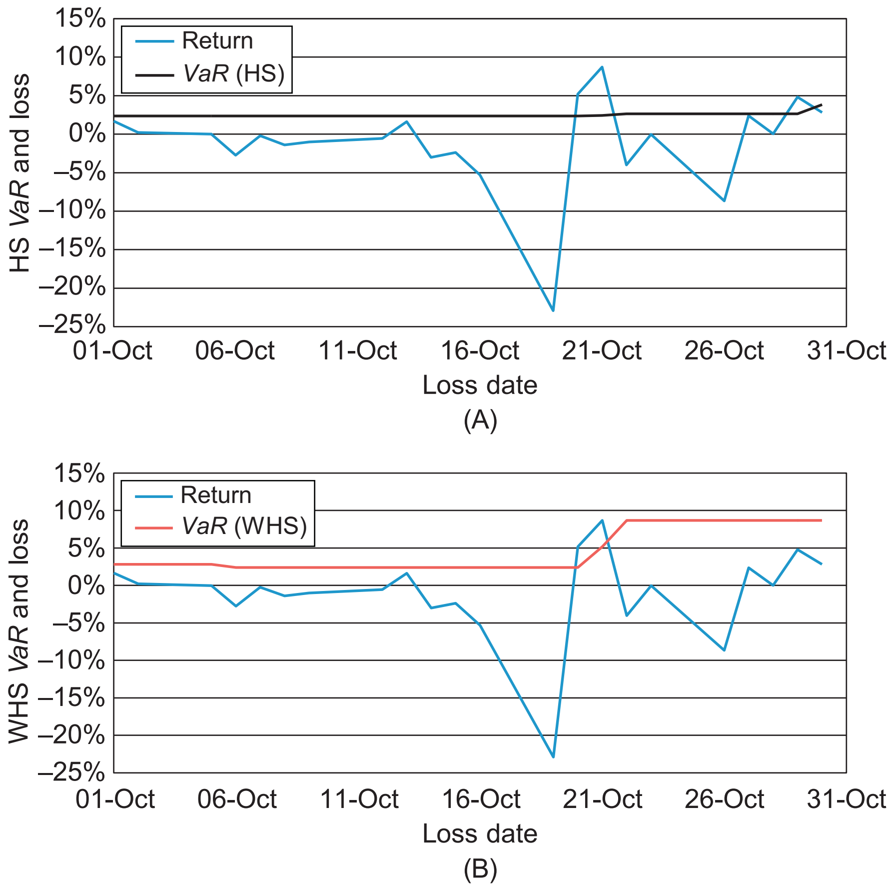


Notes: The blue line shows the daily loss in percent of $1 invested in a long position in the S&P 500 index each day during October 1987. The black line in the top panel shows the 1-day, 1% *VaR* computed using Historical Simulation with a 250-day sample. The bottom panel shows the same losses in blue and in addition the *VaR* from Weighted Historical Simulation in red.

increased, but a large gain is chalked up to the portfolio managers being clever. This is not a prudent risk management approach.

Notice that the RiskMetrics model would have picked up the increase in market variance from the crash regardless of whether the crash meant a gain or a loss to us. In

Figure 2.3 (A) Historical Simulation *VaR* and daily losses from **Short** S&P 500 position, October 1987. (B) Weighted Historical Simulation *VaR* and daily losses from **Short** S&P 500 position, October 1987.



Notes: The blue line shows the daily loss in percent of $1 invested in a short position in the S&P 500 index each day during October 1987. The black line in the top panel shows the 1-day, 1% *VaR* computed using Historical Simulation with a 250-day sample. The bottom panel shows the same losses in black and the *VaR* from Weighted Historical Simulation in red.

the RiskMetrics model, returns are squared and losses and gains are treated as having the same impact on tomorrow's variance and therefore on the portfolio risk.

Finally, a serious downside of WHS, and one it shares with the simple HS approach, is that the multiday Value-at-Risk requires a large amount of past daily return data, which is often not easy to obtain. We will study multiperiod risk modeling in Chapter 8.

4 Evidence from the 2008–2009 Crisis

The 1987 crash provides a particularly dramatic example of the problems embedded in the HS approach to *VaR* computation. The recent financial crisis involved different market dynamics than the 1987 crash but the implications for HS *VaR* are equally serious in the recent example.

Figure 2.4 shows the daily closing prices for a total return index (that is including dividends) of the S&P 500 starting in July 2008 and ending in December 2009. The index lost almost half its value between July 2008 and the market bottom in March 2009. The recovery in the index starting in March 2009 continued through the end of 2009.

HS again provides a simple way to compute *VaR*, and the red line in Figure 2.5 shows the 10-day, 1% HS *VaR*. As is standard, the 10-day *VaR* is computed from the 1-day *VaR* by simply multiplying it by $\sqrt{10}$:

$$VaR^{.01,HS}_{t+1:t+10} = -\sqrt{10} \cdot Percentile\left(\left\{R_{PF,t+1-\tau}\right\}_{\tau=1}^{m}, 1\right), \text{with } m = 250$$

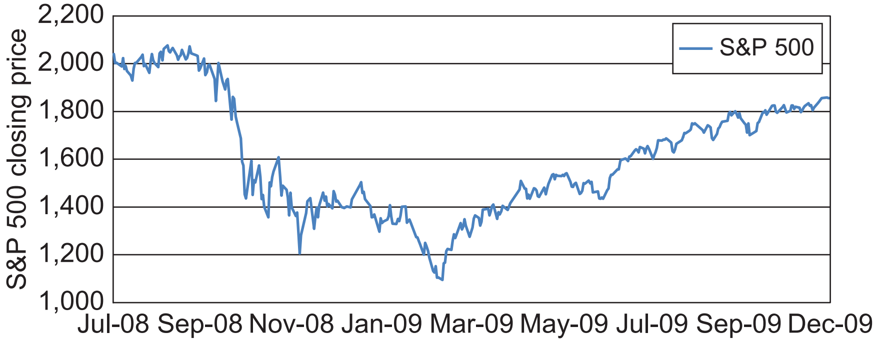
Consider now an almost equally simple alternative to HS provided by the RiskMetrics (RM) variance model discussed in Chapter 1. The blue line in Figure 2.5 shows 10-day, 1% *VaR* computed from the RiskMetrics model as follows:

$$\begin{aligned} VaR^{.01,RM}_{t+1:t+10} &= -\sqrt{10} \cdot \sigma_{t+1} \cdot \Phi^{-1}_{.01} \\ &= -\sqrt{10} \cdot \sigma_{t+1} \cdot 2.33 \end{aligned}$$

where the variance dynamics are driven by

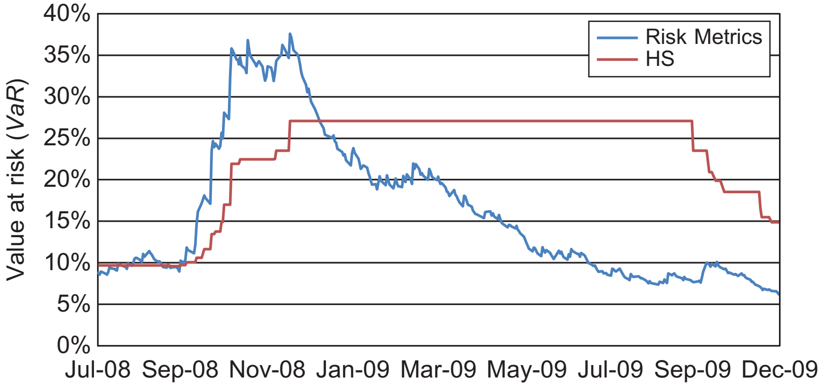
$$\sigma^2_{PF,t+1} = 0.94\sigma^2_{PF,t} + 0.06R^2_{PF,t}$$

Figure 2.4 S&P 500 total return index: 2008–2009 crisis period.



Notes: The daily closing values of the S&P 500 total return index (including dividends) are plotted from July 1, 2008 through December 31, 2009.

Figure 2.5 10-day, 1% *VaR* from Historical Simulation and RiskMetrics during the 2008–2009 crisis period.



Notes: The daily 10-day, 1% *VaR* from Historical Simulation and from RiskMetrics are plotted from July 1, 2008 through December 31, 2009.

as discussed in Chapter 1. We again simply scale the 1-day *VaR* by $\sqrt{10}$ to get the 10-day *VaR*. We have assumed a standard normal distribution for the return innovation so that the percentile is easily computed as $\Phi_{.01}^{-1} \approx -2.33$.

Notice the dramatic difference between the HS and the RM *VaR*s in Figure 2.5. The HS *VaR* rises much more slowly as the crisis gets underway in the fall of 2008 and perhaps even more strikingly, the HS *VaR* stays at its highest point for almost a year during which the volatility in the market has declined considerably. The units in Figure 2.5 refer to the least percent of capital that would be lost over the next 10 days in the 1% worst outcomes.

The upshot is that a risk management team that relies on HS *VaR* will detect the brewing crisis quite slowly and furthermore will enforce excessive caution after volatility drops in the market.

In order to put some dollar figures on this effect Figure 2.6 conducts the following experiment. Assume that each day a trader has a 10-day, 1% dollar *VaR* limit of \$100,000. Each day he or she is therefore allowed to invest

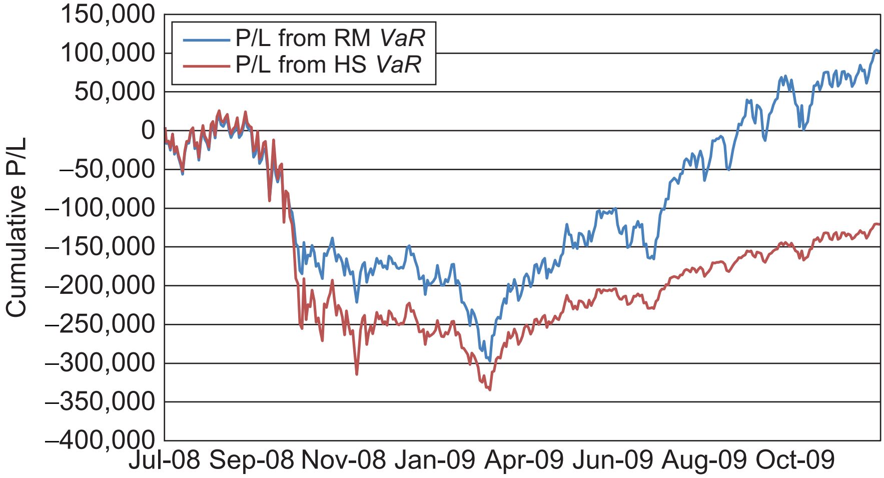
$$\$Position_{t+1} \leq \frac{\$100{,}000}{VaR_{t+1:t+10}^{.01}}$$

in the S&P 500 index.

Let us assume that the trader each day simply invests the maximum amount possible in the S&P 500, that is

$$\$Position_{t+1} = \frac{\$100,000}{VaR_{t+1:t+10}^{.01}}$$

Figure 2.6 Cumulative P/L from traders with HS and RM *VaR*s.



Notes: Using a *VaR* limit of \$100,000, a trader invests the maximum amount allowed in the S&P 500 index using RiskMetrics (blue line) and Historical Simulation (red line), respectively. The graph shows the cumulative proft and losses from the two risk models.

The red line in Figure 2.6 shows the cumulative dollar profit and loss (P/L) from a trader whose limit is based on the HS *VaR* and the blue line shows the P/L from a trader who uses the RM *VaR* model. The daily P/L is computed as

$$(P/L)_{t+1} = \$Position_{t+1}\,(S_{t+1}/S_t - 1)$$

These daily P/Ls are then cumulated across days.

The difference in performance is quite striking. The RM trader will lose less in the fall of 2008 and earn much more in 2009. The HS trader takes more dramatic losses in the fall of 2008 and is not allowed to invest sufficiently in the market in 2009 to take advantage of the run-up in the index. The HS *VaR* reacts too slowly to increases in volatility as well as to decreases in volatility. Both errors are potentially very costly.

The RM risk model is very simple—potentially too simple in several respects, which will be discussed in Chapters 4, 6, and 8. Important extensions to the simple variance dynamic assumed in RiskMetrics will be discussed in detail in Chapter 4. Recall also that we have assumed a standard normal distribution for the return innovation. This assumption is just made for convenience at this point. In Chapter 6 we will discuss ways to improve the risk model by allowing for nonnormal return innovations. Finally, we simply scaled the 1-day *VaR* by $\sqrt{10}$ to get the 10-day *VaR*. This simple rule is an approximation that is often not accurate. It will be discussed in detail in Chapter 8.

5 The True Probability of Breaching the HS *VaR*

The 1987 and the 2008–2009 examples indicated the problems inherent in the HS approach but they could be dismissed as being just examples and of course not randomly chosen periods. In order to get beyond these concerns we now conduct the following purely artificial but revealing experiment.

Assume that the S&P 500 market returns are generated by a time series process with dynamic volatility and normal innovations. In reality of course they are not but if we make this radically simplifying assumption then we are able to compute how wrong the HS *VaR* can get. To be specific, assume that innovation to S&P 500 returns each day is drawn from the normal distribution with mean zero and variance equal to $\sigma^2_{PF,t+1}$,which is generated by a GARCH type variance process that we will introduce in Chapter 4. We can write

$$R_{PF,t+1} = \sigma_{PF,t+1} z_{PF,t+1}, \quad \text{with } z_{t+1} \sim i.i.d.\ N(0,1)$$

If we simulate 1,250 return observations from this process, then starting on day 251 we can, on each day, compute the 1-day, 1% *VaR* using Historical Simulation. Because we know how the returns were created, we can, on each day, compute the true probability that we will observe a loss larger than the HS *VaR* we have computed. We call this the probability of a *VaR* breach. It is computed as

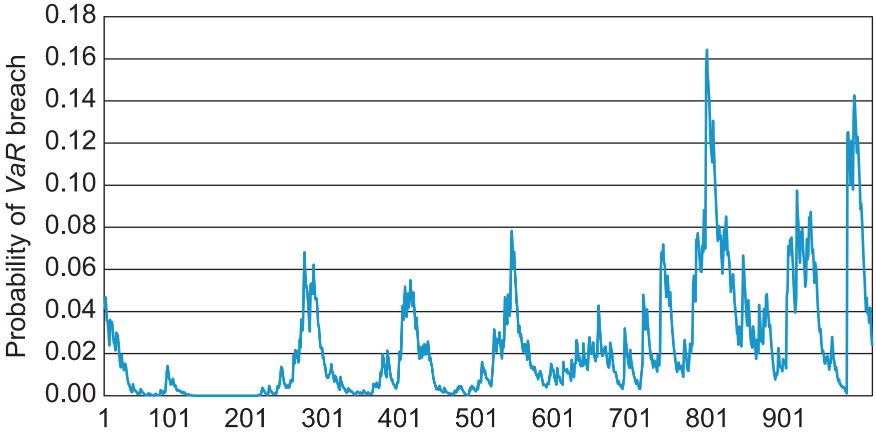
$$\begin{aligned}
\Pr\left(R_{PF,t+1} < -VaR^{.01,HS}_{t+1}\right) &= \Pr\left(R_{PF,t+1}/\sigma_{PF,t+1} < -VaR^{.01,HS}_{t+1}/\sigma_{PF,t+1}\right) \\
&= \Pr\left(z_{PF,t+1} < -VaR^{.01,HS}_{t+1}/\sigma_{PF,t+1}\right) \\
&= \Phi\left(-VaR^{.01,HS}_{t+1}/\sigma_{PF,t+1}\right)
\end{aligned}$$

where Φ is again the cumulative density function for a standard normal random variable. Figure 2.7 shows this probability over the 1,000 simulated return days.

If the HS *VaR* model had been accurate then this plot should show a roughly flat line at 1%. Instead we see numbers as high as 16%, which happens when volatility is high, and numbers very close to 0%, which happens when volatility is low. The HS *VaR* will tend to overestimate risk when the true market volatility is low, which will generate a low probability of a *VaR* breach in Figure 2.7. Conversely, and more crucially, HS will underestimate risk when true volatility is high in which case the *VaR* breach volatility will be high. The HS approach, which is supposed to deliver a 1% *VaR*, sometimes delivers a 16% *VaR*, which means that there is roughly a 1 in 6 chance of getting a loss worse than the HS *VaR*, when there is supposed to be only a 1 in 100 chance. The upshot is that HS *VaR* may be roughly correct on average (the average of the probabilities in Figure 2.7 is 2.3%) but the HS *VaR* is much too low when volatility is high and the HS *VaR* is too high when volatility is low.

This example has used Monte Carlo simulation to generate artificial returns. We will study the details of Monte Carlo simulation in Chapter 8.

Figure 2.7 Actual probability of losing more than the 1% HS *VaR* when returns have dynamic variance.



Notes: The figure shows the probability of getting a return worse than the *VaR* when the return is simulated from a model with dynamic variance and the *VaR* is computed using Historical Simulation.

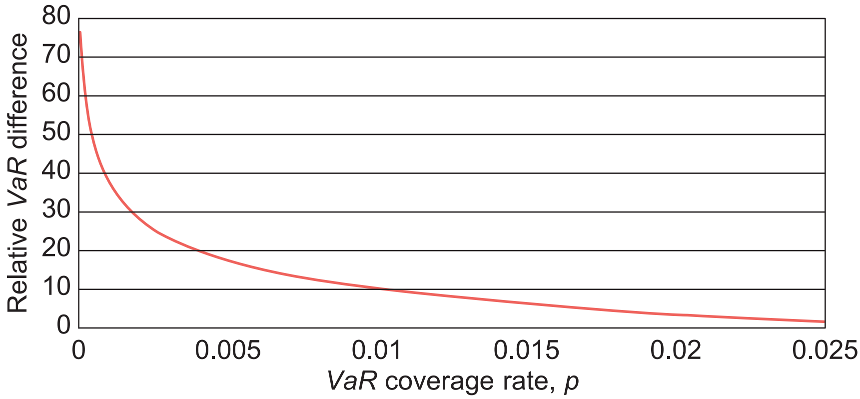
6 *VaR* with Extreme Coverage Rates

The most complete picture of risk is no doubt provided by reporting the entire shape of the tail of the distribution of losses beyond the *VaR*. The tail of the portfolio return distribution, when modeled correctly, tells the risk manager everything about the future losses. Reporting the entire tail of the return distribution corresponds to reporting *VaR*s for many different coverage rates, say p ranging from 0.01% to 2.5% in increments. Note that when using HS with a 250-day sample it is not even possible to compute the *VaR* when $p < 1/250 = 0.4\%$.

Figure 2.8 illustrates the relative difference between a *VaR* from a nonnormal distribution (with an excess kurtosis of 3) and a *VaR* from a normal distribution as a function of the *VaR* probability, p. Notice that as p gets close to zero (the smallest p in the figure is 0.0001, which is 0.01%) the nonnormal *VaR* gets much larger than the normal *VaR*. Strikingly, when $p = 0.025$ (i.e., 2.5%) there is almost no difference between the two *VaR*s even though the underlying distributions are actually quite different as the *VaRs* with extreme ps show. Relying on *VaR* with large p is dangerous because extreme risks are hidden. Chapter 6 will detail the important task of modeling nonnormality in the return distribution.

The popularity of *VaR* as a risk measurement tool is due to its simple interpretation: "What's the loss so that only $100p\%$ of potential losses tomorrow will be worse?" However, reporting the *VaR* for several values of p, where p is small, should be given serious consideration in risk reporting as it maps out the tail of the loss distribution.

Figure 2.8 Relative difference between nonnormal (excess kurtosis = 3) and normal *VaR*.



Notes: The figure plots [*VaR*(NonN)-*VaR*(N)]/*VaR*[N] where "N" denotes normal distribution. The *VaR* difference is shown as a function of the VaR coverage rate, *p*.

7 Expected Shortfall

We previously discussed a key shortcoming of *VaR*, namely that it is concerned only with the percentage of losses that exceed the *VaR* and not the magnitude of these losses. The magnitude, however, should be of serious concern to the risk manager. Extremely large losses are of course much more likely to cause financial distress, such as bankruptcy, than are moderately large losses; therefore we want to consider a risk measure that accounts for the magnitude of large losses as well as their probability of occurring.

The challenge is to come up with a portfolio risk measure that retains the simplicity of the *VaR*, but conveys information regarding the shape of the tail. Expected Shortfall (*ES*), or TailVaR as it is sometimes called, is one way to do this.

Mathematically *ES* is defined as

$$ES^p_{t+1} = -E_t\left[R_{PF,t+1} | R_{PF,t+1} < -VaR^p_{t+1}\right]$$

where the negative signs in front of the expectation and the *VaR* are needed because the *ES* and the *VaR* are defined as positive numbers. The Expected Shortfall tells us the expected value of tomorrow's loss, conditional on it being worse than the *VaR*.

The distribution tail gives us information on the range of possible extreme losses and the probability associated with each outcome. The Expected Shortfall measure aggregates this information into a single number by computing the average of the tail outcomes weighted by their probabilities. So where *VaR* tells us the loss so that only 1% of potential losses will be worse, the *ES* tells us the expected loss given that we actually get a loss from the 1% tail. So while we are not conveying all the information

in the shape of the tail when using *ES*, the key is that the shape of the tail beyond the *VaR* measure is now important for determining the risk number.

To gain more insight into the *ES* as a risk measure, let's first consider the normal distribution. In order to compute *ES* we need the distribution of a normal variable conditional on it being below the *VaR*. The truncated standard normal distribution is defined from the standard normal distribution as

$$\phi_{Tr}(z|z \le Tr) = \frac{\phi(z)}{\Phi(Tr)} \quad \text{with } E[z|z \le Tr] = -\frac{\phi(Tr)}{\Phi(Tr)}$$

where $\phi(\bullet)$ denotes the density function and $\Phi(\bullet)$ the cumulative density function of the standard normal distribution.

Recall that $R_{PF,t+1} = \sigma_{PF,t+1} z_{PF,t+1}$. In the normal distribution case *ES* can therefore be derived as

$$\begin{aligned} ES^p_{t+1} &= -E_t\left[R_{PF,t+1} | R_{PF,t+1} \le -VaR^p_{t+1}\right] \\ &= -\sigma_{PF,t+1} E_t\left[z_{PF,t+1} | z_{PF,t+1} \le -VaR^p_{t+1}/\sigma_{PF,t+1}\right] \\ &= \sigma_{PF,t+1} \frac{\phi\left(-VaR^p_{t+1}/\sigma_{PF,t+1}\right)}{\Phi\left(-VaR^p_{t+1}/\sigma_{PF,t+1}\right)} \end{aligned}$$

Of course, in the normal case we also know that

$$VaR^p_{t+1} = -\sigma_{PF,t+1}\Phi^{-1}_p$$

Thus, we have

$$ES^p_{t+1} = \sigma_{PF,t+1} \frac{\phi\left(\Phi^{-1}_p\right)}{p}$$

which has a structure very similar to the *VaR* measure.

The relative difference between *ES* and *VaR* is

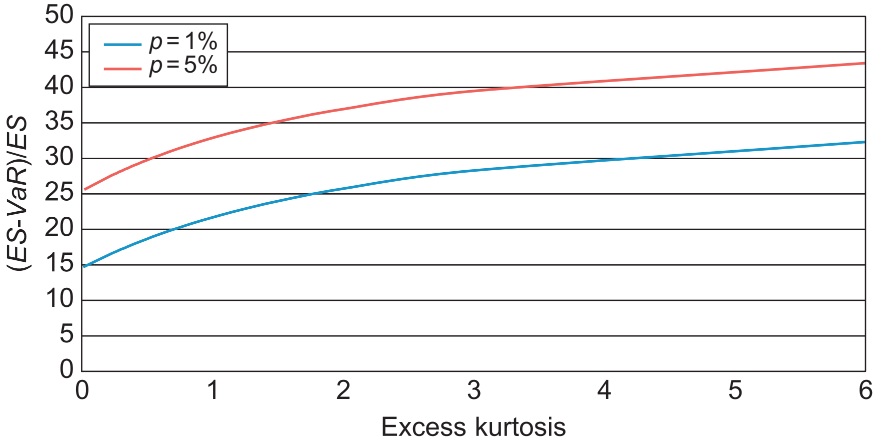
$$\frac{ES^p_{t+1} - VaR^p_{t+1}}{VaR^p_{t+1}} = -\frac{\phi\left(\Phi^{-1}_p\right)}{p\Phi^{-1}_p} - 1$$

When, for example, $p = 0.01$, we have $\Phi^{-1}_p \approx -2.33$, and the relative difference is then

$$\frac{ES^{.01}_{t+1} - VaR^{.01}_{t+1}}{VaR^{.01}_{t+1}} \approx -\frac{(2\pi)^{-1/2}\exp(-(-2.33)^2/2)}{.01(-2.33)} - 1 \approx 15\%$$

In the normal case, we can show that as the *VaR* coverage rate *p* gets close to zero, the ratio of the *ES* to the *VaR* goes to 1.

Figure 2.9 *ES* versus *VaR* as a function of kurtosis.



Notes: The figure shows (*ES*-*VaR*)/*VaR* in percent as a function of the excess kurtosis of the underlying portfolio return distribution. The blue line uses a 1% *VaR* and the red line uses a 5% *VaR*.

From this it would seem that it really doesn't matter much whether the risk manager uses *VaR* or *ES* as a risk measure. The difference is only 15% when the *VaR* coverage rate is 1%. Recall, however, that we argued in Chapter 1 that normal distributions fit asset return data poorly, particularly in the tail. So what happens to the $(ES - VaR)/VaR$ ratio when we look at nonnormal distributions?

Figure 2.9 considers a fat-tailed distribution where the degree of fatness in the tail is captured by excess kurtosis as defined in Chapter 1: the higher the excess kurtosis the fatter the distribution tail. The blue line in Figure 2.9 covers the case where $p = 1\%$ and the red line shows $p = 5\%$.

The blue line shows that when excess kurtosis is zero we get that the relative difference between the *ES* and *VaR* is 15%, which matches the preceding computation for the normal distribution. The blue line in Figure 2.9 also shows that for moderately large values of excess kurtosis, the relative difference between *ES* and *VaR* is above 30%.

Comparing the red line with the blue line in Figure 2.9 it is clear that the relative difference between *VaR* and *ES* is larger when *p* is larger and thus further from zero. When *p* is close to zero *VaR* and *ES* will both capture the fat tails in the distribution. When *p* is far from zero, only the *ES* will capture the fat tails in the return distribution. When using *VaR* and a large *p* the dangerous large losses will be hidden from view. Generally an *ES*-based risk measure will be better able to capture the fact that a portfolio with large kurtosis (or negative skewness) is more risky than a portfolio with low kurtosis.

Risk managers who rely on Historical Simulation often report *VaR* with relatively large *p* because they are worried about basing the *VaR* estimate on too few

observations. Figure 2.9 indirectly shows that this argument has a serious downside: The larger the p the more likely it is that extreme risks (evident in *ES*) in the portfolio return distribution will go unnoticed in the *VaR*. The *ES* risk measure will capture such extreme risks. The *VaR* will not.

8 Summary

VaR is the most popular risk measure in use and HS is the most often used methodology to compute *VaR*. This chapter has argued that *VaR* as commonly reported has some shortcomings and that using HS to compute *VaR* has serious problems as well.

We need instead to use risk measures that capture the degree of fatness in the tail of the return distribution, and we need risk models that properly account for the dynamics in variance and models that can be used across different return horizons.

Going forward the goal will be to develop risk models with the following characteristics:

- The model is a fully specified statistical process that can be estimated on daily returns.
- The model can be estimated and implemented for portfolios with a large number of assets.
- *VaR* and *ES* can be easily computed for any prespecified level of confidence, p, and for any horizon of interest, K.
- *VaR* and *ES* are dynamic reflecting current market conditions.

In order to deliver accurate risk predictions, the model should reflect the following stylized facts of daily asset returns discussed in Chapter 1:

- The expected daily returns have little or no predictability.
- The variance of daily returns greatly exceeds the mean.
- The variance of daily returns is predictable.
- Daily returns are not normally distributed.
- Even after standardizing daily returns by a dynamic variance model, the standardized daily returns are not normally distributed.
- Positive and negative returns of the same magnitude may have different impacts on the variance.
- Correlations between assets appear to be time-varying.
- As the investment horizon increases, the return data distribution approaches the normal distribution.

Further Resources

Useful overviews of the various approaches to *VaR* calculation can be found in Duffie and Pan (1997), Engle and Manganelli (2004a), Jorion (2006), and Christoffersen (2009).

Dowd and Blake (2006) discuss the use of *VaR*-like measures in the insurance industry. Danielsson (2002) warns against using risk models estimated on asset return data from calm markets.

Bodoukh et al. (1998) introduced the Weighted Historical Simulation approach. They found that it compares favorably with both the HS approach and the RiskMetrics model. Figures 2.2 and 2.3 are based on Pritsker (2006).

Engle and Manganelli (2004b) suggest an interesting alternative method (not discussed in this chapter) for *VaR* calculation based on conditional quantile regression.

Artzner et al. (1999) define the concept of a coherent risk measure and showed that Expected Shortfall (*ES*) is coherent whereas *VaR* is not. Inui and Kijima (2005) provide additional theoretical arguments for the use of *ES*. Taylor (2008) provides econometric tools for *ES* computation.

Studying dynamic portfolio management based on *ES* and *VaR*, Basak and Shapiro (2001) found that when a large loss does occur, *ES* risk management leads to lower losses than *VaR* risk management. Cuoco et al. (2008) argued instead that *VaR* and *ES* risk management lead to equivalent results as long as the *VaR* and *ES* risk measures are recalculated often. Both Basak and Shapiro (2001) and Cuoco et al. (2008) assumed that returns are normally distributed. Yamai and Yoshiba (2005) compare *VaR* and *ES* from a practical perspective. Berkowitz and O'Brien (2002) and Alexander and Baptista (2006) look at VaR from a regulatory perspective.

References

Alexander, G.J., Baptista, A.M., 2006. Does the Basle Capital Accord reduce bank fragility? An assessment of the Value-at-Risk approach. J. Monet. Econom. 53, 1631–1660.

Artzner, P., Delbaen, F., Eber, J., Heath, D., 1999. Coherent measures of risk. Math. Finance 9, 203–228.

Basak, S., Shapiro, A., 2001. Value-at-risk-based risk management: Optimal policies and asset prices. Rev. Financial Stud. 14, 371–405.

Berkowitz, J., O'Brien, J., 2002. How accurate are Value-at-Risk models at commercial banks? J. Finance 57, 1093–1111.

Bodoukh, J., Richardson, M., Whitelaw, R., 1998. The best of both worlds. Risk 11, 64–67.

Christoffersen, P., 2009. Value-at-Risk models. In: Mikosch, T., Kreib, J.-P., Davis, R.A., Andersen, T.G., (Eds.), Handbook of Financial Time Series, Springer Verlag, Berlin, pp. 753–766.

Cuoco, D., He, H., Issaenko, S., 2008. Optimal dynamic trading strategies with risk limits. Oper. Res. 56, 358–368.

Danielsson, J., 2002. The emperor has no clothes: Limits to risk modelling. J. Bank. Finance 26, 1273–1296.

Dowd, K., Blake, D., 2006. After VaR: The theory, estimation, and insurance applications of quantile-based risk measures. J. Risk Insur. 73, 193–229.
Duffie, D., Pan, J., 1997. An overview of value at risk. J. Derivatives 4, 7–49.
Engle, R., Manganelli, S., 2004a. A comparison of value at risk models in finance. In: Giorgio Szego, (Ed.), Risk Measures for the 21st Century, Wiley Finance, West Sussex England.
Engle, R., Manganelli, S., 2004b. CAViaR: Conditional value at risk by quantile regression. J. Bus. Econom. Stat. 22, 367–381.
Inui, K., Kijima, M., 2005. On the significance of expected shortfall as a coherent risk measure. J. Bank. Finance 29, 853–864.
Jorion, P., 2006. Value-at-Risk: The New Benchmark for Managing Financial Risk, third ed. McGraw-Hill, New York, NY.
Pritsker, M., 2006. The hidden dangers of historical simulation. J. Bank. Finance 30, 561–582.
Taylor, J.W., 2008. Estimating value at risk and expected shortfall using expectiles. J. Financial Econom. 6, 231–252.
Yamai, Y., Yoshiba, T., 2005. Value-at-Risk versus expected shortfall: A practical perspective. J. Bank. Finance 29, 997–1015.

Empirical Exercises

Open the Chapter2data.xlsx file on the web site. Use sheet 1 for questions 1 and 2, and sheet 2 for questions 3 and 4.

1. Assume you are long \$1 of the S&P 500 index on each day. Calculate the 1-day, 1% *VaR*s on each day in October 1987 using Historical Simulation. Use a 250-day moving window. Plot the *VaR* and the losses. Repeat the exercise assuming you are short \$1 each day. Plot the *VaR* and the losses again. Compare with Figures 2.2 and 2.3.
2. Assume you are long \$1 of the S&P 500 index on each day. Calculate the 1-day, 1% *VaR*s on each day in October 1987 using Weighted Historical Simulation. You can ignore the linear interpolation part of WHS. Use a weighting parameter of $\eta = 0.99$ in WHS. Use a 250-day moving window. (*Excel hint*: Sort the returns along with their weights by selecting both columns in Excel and sorting by returns.) Repeat the exercise assuming you are short \$1 each day. Plot the *VaR* and the losses again. Compare with Figures 2.2 and 2.3.
3. For each day from July 1, 2008 through December 31, 2009, calculate the 10-day, 1% *VaR*s using the following methods: (a) RiskMetrics, that is, normal distribution with an exponential smoother on variance using the weight, $\lambda = 0.94$; and (b) Historical Simulation. Use a 250-day moving sample. Compute the 10-day *VaR*s from the 1-day *VaR*s by just multiplying by square root of 10. Plot the *VaR*s.
4. Reconstruct the P/Ls in Figure 2.6.

The answers to these exercises can be found in the Chapter2Results.xlsx file on the companion website.

For more information see the companion site at
http://www.elsevierdirect.com/companions/9780123744487

3 A Primer on Financial Time Series Analysis

1 Chapter Overview

This chapter serves two purposes: First, it gives a very brief refresher on the basic concepts in probability and statistics, and introduces the bivariate linear regression model. Second, it gives an introduction to time series analysis with a focus on the models most relevant for financial risk management. The chapter can be skipped by readers who have recently taken a course in time series analysis or in financial econometrics.

The material in the chapter is organized in the following four sections:

1. Probability Distributions and Moments
2. The Linear Model
3. Univariate Time Series Models
4. Multivariate Time Series Models

The chapter thus tries to cover a broad range of material that really would take several books to do justice. The section “Further Resources” at the end of the chapter therefore suggests books that can be consulted for readers who need to build a stronger foundation in statistics and econometrics and also for readers who are curious to tackle more advanced topics in time series analysis.

An important goal of the financial time series analysis part of the chapter is to ensure that the reader avoids some common pitfalls encountered by risk managers working with time series data such as prices and returns. These pitfalls can be summarized as

- Spurious detection of mean-reversion; that is, erroneously finding that a variable is mean-reverting when it is truly a random walk
- Spurious regression; that is, erroneously finding that a variable x is significant in a regression of y on x
- Spurious detection of causality; that is, erroneously finding that the current value of x causes (helps determine) future values of y when in reality it cannot

Before proceeding to these important topics in financial time series analysis we first provide a quick refresher on basic probability and statistics.

Elements of Financial Risk Management. DOI: 10.1016/B978-0-12-374448-7.00003-8

2 Probability Distributions and Moments

The probability distribution of a discrete random variable, x, describes the probability of each possible outcome of x. Even if an asset price in reality can only take on discrete values (for example \$14.55) and not a continuum of values (for example \$14.55555.....) we usually use continuous densities rather than discrete distributions to describe probability of various outcomes. Continuous probability densities are more analytically tractable and they approximate well the discrete probability distributions relevant for risk management.

2.1 Univariate Probability Distributions

Let the function $F(x)$ denote the cumulative probability distribution function of the random variable x so that the probability of x being less than the value a is given by

$$\Pr(x < a) = F(a)$$

Let $f(x)$ be the probability density of x and assume that x is defined from $-\infty$ to $+\infty$. The probability of obtaining a value of x less that a can be had from the density via the integral

$$\Pr(x < a) = \int_{-\infty}^{a} f(x)dx = F(a)$$

so that $f(x) = \frac{\partial F(x)}{x}$. We also have that

$$\Pr(x < +\infty) = \int_{-\infty}^{+\infty} f(x)dx = 1, \text{ and}$$
$$\Pr(x = a) = 0$$

Because the density is continuous the probability of obtaining any particular value a is zero. The probability of obtaining a value in an interval between b and a is

$$\Pr(b < x < a) = \int_{b}^{a} f(x)dx = F(a) - F(b), \quad \text{where } b < a$$

The expected value or mean of x captures the average outcome of a draw from the distribution and it is defined as the probability weighted average of x

$$E[x] = \int_{-\infty}^{\infty} xf(x)dx$$

The basic rules of integration and the property that $\int_{-\infty}^{+\infty} f(x)dx = 1$ provides useful results for manipulating expectations, for example

$$E[a+bx] = \int_{-\infty}^{\infty} (a+bx)f(x)dx = a + b\int_{-\infty}^{\infty} xf(x)dx = a + bE[x]$$

where a and b are constants.

Variance is a measure of the expected variation of variable around its mean. It is defined by

$$Var[x] = E\left[(x-E[x])^2\right] = \int_{-\infty}^{\infty} (x-E[x])^2 f(x)dx$$

Note that

$$Var[x] = E\left[(x-E[x])^2\right] = E\left[x^2 + E[x]^2 - 2xE[x]\right] = E\left[x^2\right] - E[x]^2$$

which follows from $E[E[x]] = E[x]$. From this we have that

$$\begin{aligned} Var\left[a+bx\right] &= E\left[(a+bx)^2\right] - E[(a+bx)]^2 \\ &= b^2E[x^2] - b^2E[x]^2 \\ &= b^2 Var[x] \end{aligned}$$

The standard deviation is defined as the square root of the variance. In risk management, volatility is often used as a generic term for either variance or standard deviation.

From this note, if we define a variable $y = a + bx$ and if the mean of x is zero and the variance of x is one then

$$\begin{aligned} E[y] &= a \\ Var[y] &= b^2 \end{aligned}$$

This is useful for creating variables with the desired mean and variance.

Mean and variance are the first two central moments. The third and fourth central moments, also known as skewness and kurtosis, are defined by:

$$\begin{aligned} Skew[x] &= \frac{\int_{-\infty}^{\infty} (x-E[x])^3 f(x)dx}{Var[x]^{3/2}} \\ Kurt[x] &= \frac{\int_{-\infty}^{\infty} (x-E[x])^4 f(x)dx}{Var[x]^2} \end{aligned}$$

Note that by subtracting $E[x]$ before taking powers and by dividing skewness by $Var[x]^{3/2}$ and kurtosis by $Var[x]^2$ we ensure that

$$Skew[a+bx] = Skew[x]$$
$$Kurt[a+bx] = Kurt[x]$$

and we therefore say that skewness and kurtosis are location and scale invariant.

As an example consider the normal distribution with parameters μ and σ^2. It is defined by

$$f\left(x;\mu,\sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$$

The normal distribution has the first four moments

$$E[x] = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \mu$$

$$Var[x] = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \sigma^2$$

$$Skew[x] = \frac{1}{\sigma^3} \int_{-\infty}^{\infty} (x-\mu)^3 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right) dx = 0$$

$$Kurt[x] = \frac{1}{\sigma^4} \int_{-\infty}^{\infty} (x-\mu)^4 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right) dx = 3$$

2.2 Bivariate Distributions

When considering two random variables x and y we can define the bivariate density $f(x,y)$ so that

$$\Pr(a < x < b, c < y < d) = \int_c^d \int_a^b f(x,y) dx dy$$

Covariance is the most common measure of linear dependence between two variables. It is defined by

$$Cov[x,y] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E[x])(y - E[y]) f(x,y) dx dy$$

From the properties of integration we have the following convenient result:

$$Cov[a+bx, c+dy] = bdCov[x,y]$$

so that the covariance depends on the magnitude of x and y but not on their means. Note also from the definition of covariance that

$$Cov[x,x] = \int_{-\infty}^{\infty} (x - E[x])^2 f(x)\,dx = Var[x]$$

From the covariance and variance definitions we can define correlation by

$$Corr[x,y] = \frac{Cov[x,y]}{\sqrt{Var[x]\,Var[y]}}$$

Notice that the correlation between x and y does not depend on the magnitude of x and y. We have

$$Corr[a+bx, c+dy] = \frac{bdCov[x,y]}{\sqrt{b^2 Var[x]\, d^2 Var[y]}} = \frac{Cov[x,y]}{\sqrt{Var[x]\,Var[y]}} = Corr[x,y]$$

A perfect positive linear relationship between x and y would exist if $y = a + bx$, in which case

$$Corr[x,y] = \frac{Cov[x, a+bx]}{\sqrt{Var[x]\,Var[a+bx]}} = \frac{bVar[x]}{bVar[x]} = 1$$

A perfect negative linear relationship between x and y exists if $y = a - bx$, in which case

$$Corr[x,y] = \frac{Cov[x, a-bx]}{\sqrt{Var[x]\,Var[a-bx]}} = \frac{-bVar[x]}{bVar[x]} = -1$$

This suggests that correlation is bounded between -1 and $+1$, which is indeed the case. This fact is convenient when interpreting a given correlation value.

2.3 Conditional Distributions

Risk managers often want to describe a variable y using information on another variable x. From the joint distribution of x and y we can denote the conditional distribution of y given x, $f(y|x)$. It must be the case that

$$f(x,y) = f(y|x)f(x)$$

which indirectly defines the conditional distribution as

$$f(y|x) = \frac{f(x,y)}{f(x)}$$

This definition can be used to define the conditional mean and variance

$$E[y|x] = \int_{-\infty}^{\infty} yf(y|x)dy = \int_{-\infty}^{\infty} y\frac{f(x,y)}{f(x)}dy$$

$$Var[y|x] = \int_{-\infty}^{\infty} (y - E[y|x])^2 \frac{f(x,y)}{f(x)}dy$$

Note that these conditional moments are functions of x but not of y.

If x and y are independent then $f(y|x) = f(y)$ and so $f(x,y) = f(x)f(y)$ and we have that the conditional moments

$$E[y|x] = \int_{-\infty}^{\infty} y\frac{f(x)f(y)}{f(x)}dy = \int_{-\infty}^{\infty} yf(y)dy = E[y]$$

$$Var[y|x] = \int_{-\infty}^{\infty} (y - E[y])^2 \frac{f(x)f(y)}{f(x)}dy = Var[y]$$

equal the corresponding unconditional moments.

2.4 Sample Moments

We now introduce the standard methods for estimating the moments introduced earlier.

Consider a sample of T observations of the variable x, namely $\{x_1, x_2, \ldots, x_T\}$. We can estimate the mean using the sample average

$$\widehat{E}[x] = \bar{x} = \tfrac{1}{T}\sum_{t=1}^{T} x_t$$

and we can estimate the variance using the sample average of squared deviations from the average

$$\widehat{Var}[x] = \tfrac{1}{T}\sum_{t=1}^{T} (x_t - \bar{x})^2$$

Sometimes the sample variance uses $\frac{1}{T-1}$ instead of $\frac{1}{T}$ but unless T is very small then the difference can be ignored.

Similarly skewness and kurtosis can be estimated by

$$\widehat{Skew}[x] = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^3 / \widehat{Var}[x]^{3/2}$$
$$\widehat{Kurt}[x] = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^4 / \widehat{Var}[x]^2$$

The sample covariances between two random variables can be estimated via

$$\widehat{Cov}[x,y] = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})$$

and the sample correlation between two random variables, x and y, is calculated as

$$\widehat{\rho}_{x,y} = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^{T}(x_t - \bar{x})^2 \sum_{t=1}^{T}(y_t - \bar{y})^2}}$$

3 The Linear Model

Risk managers often rely on linear models of the type

$$y = a + bx + \varepsilon$$

where $E[\varepsilon] = 0$ and x and ε are assumed to be independent or sometimes just uncorrelated. If we know the value of x then we can use the linear model to predict y via the conditional expectation of y given x

$$E[y|x] = a + bE[x|x] + E[\varepsilon|x] = a + bx$$

In the linear model the unconditional expectations of x and y are linked via

$$E[y] = a + bE[x] + E[\varepsilon] = a + bE[x]$$

so that

$$a = E[y] - bE[x]$$

We also have that

$$Cov[y,x] = Cov[a + bx + \varepsilon, x] = bCov[x,x] = bVar[x]$$

so that

$$b = \frac{Cov[y,x]}{Var[x]}$$

In the linear model the variances of x and y are linked via

$$Var[y] = b^2 Var[x] + Var[\varepsilon]$$

Consider observation t in the linear model

$$y_t = a + bx_t + \varepsilon_t$$

If we have a sample of T observations then we can estimate

$$\hat{b} = \frac{\widehat{Cov}[x,y]}{\widehat{Var}[x]} = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{T}(x_t - \bar{x})^2}$$

and

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

In the more general linear model with J different x-variables we have

$$y_t = a + \sum_{j=1}^{J} b_j x_{j,t} + \varepsilon_t$$

Minimizing the sum of squared errors, $\sum_{t=1}^{T}\varepsilon_t^2$ provides the ordinary least square (OLS) estimate of b:

$$\hat{b} = \arg\min \sum_{t=1}^{T} \varepsilon_t^2 = \arg\min \sum_{t=1}^{T} \left(y_t - a - \sum_{j=1}^{J} b_j x_{j,t} \right)^2$$

The solution to this optimization problem is a linear function of y and x, which makes OLS estimation very easy to perform; thus it is built in to most common quantitative software packages such as Excel, where the OLS estimation function is called LINEST.

3.1 The Importance of Data Plots

While the linear model is useful in many cases, an apparent linear relationship between two variables can be deceiving. Consider the four (artificial) data sets in Table 3.1,

Table 3.1 Anscombe's quartet

	I		II		III		IV	
	x	y	x	y	x	y	x	y
	10	8.04	10	9.14	10	7.46	8	6.58
	8	6.95	8	8.14	8	6.77	8	5.76
	13	7.58	13	8.74	13	12.74	8	7.71
	9	8.81	9	8.77	9	7.11	8	8.84
	11	8.33	11	9.26	11	7.81	8	8.47
	14	9.96	14	8.1	14	8.84	8	7.04
	6	7.24	6	6.13	6	6.08	8	5.25
	4	4.26	4	3.1	4	5.39	19	12.5
	12	10.84	12	9.13	12	8.15	8	5.56
	7	4.82	7	7.26	7	6.42	8	7.91
	5	5.68	5	4.74	5	5.73	8	6.89
Moments								
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	11.0	4.1	11.0	4.1	11.0	4.1	11.0	4.1
Correlation	0.82		0.82		0.82		0.82	
Regression								
a	3.00		3.00		3.00		3.00	
b	0.50		0.50		0.50		0.50	

Notes: The table contains the four bivariate data sets in Anscombe's quartet. Below each of the eight variables we report the mean and the variance. We also report the correlation between x and y in each of the four data sets. The parameter a denotes the constant and b denotes the slope from the regression of y on x.

which are known as Anscombe's quartet, named after their creator. All four data sets have 11 observations.

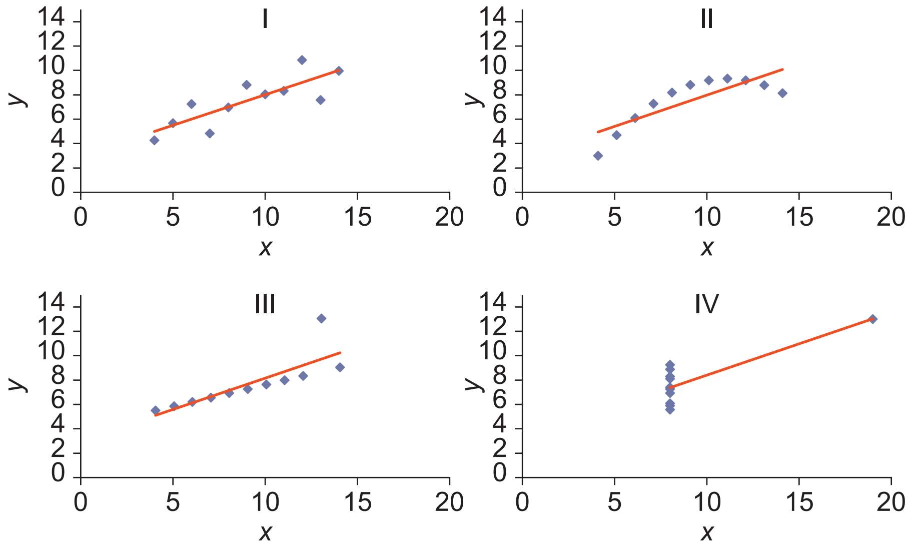
Consider now the moments of the data included at the bottom of the observations in Table 3.1. While the observations in the four data sets are clearly different from each other, the mean and variance of the x and y variables is exactly the same across the four data sets. Furthermore, the correlation between x and y are also the same across the four pairs of variables. Finally, the last two rows of Table 3.1 show that when regressing y on x using the linear model

$$y_t = a + bx_t + \varepsilon_t$$

we get the parameter estimates $a = 3$ and $b = 0.5$ in all the four cases. This data has clearly been reverse engineered by Anscombe to produce such striking results.

Figure 3.1 scatter plots y against x in the four data sets with the regression line included in each case. Figure 3.1 is clearly much more revealing than the moments and the regression results.

Figure 3.1 Scatter plot of Anscombe's four data sets with regression lines.



Notes: For each of the four data sets in Anscombe's quartet we scatter plot the variables and also report the regression line from fitting *y* on *x*.

We conclude that moments and regressions can be useful for summarizing variables and relationships between them but whenever possible it is crucial to complement the analysis with figures. When plotting your data you may discover:

- A genuine linear relationship as in the top-left panel of Figure 3.1
- A genuine nonlinear relationship as in the top-right panel
- A biased estimate of the slope driven by an outlier observation as in the bottom-left panel
- A trivial relationship, which appears as a linear relationship again due to an outlier as in the bottom-right panel of Figure 3.1

Remember: Always plot your variables before beginning a statistical analysis of them.

4 Univariate Time Series Models

Univariate time series analysis studies the behavior of a single random variable observed over time. Risk managers are interested in how prices and risk factors move over time; therefore time series models are useful for risk managers. Forecasting the future values of a variable using past and current observations on the same variable is a key topic in univariate time series analysis.

4.1 Autocorrelation

Correlation measures the linear dependence between two variables and autocorrelation measures the linear dependence between the current value of a time series variable and the past value of the same variable. Autocorrelation is a crucial tool for detecting linear dynamics in time series analysis.

The autocorrelation for lag τ is defined as

$$\rho_\tau \equiv Corr\left[R_t, R_{t-\tau}\right] = \frac{Cov\left[R_t, R_{t-\tau}\right]}{\sqrt{Var\left[R_t\right] Var\left[R_{t-\tau}\right]}} = \frac{Cov\left[R_t, R_{t-\tau}\right]}{Var\left[R_t\right]}$$

so that it captures the linear relationship between today's value and the value τ days ago.

Consider a data set on an asset return, $\{R_1, R_2, \ldots, R_T\}$. The sample autocorrelation at lag τ measures the linear dependence between today's return, R_t, and the return τ days ago, $R_{t-\tau}$. Using the autocorrelation definition, we can write the sample autocorrelation as

$$\widehat{\rho}_\tau = \frac{\frac{1}{T-\tau}\sum_{t=\tau+1}^{T}\left(R_t - \overline{R}\right)\left(R_{t-\tau} - \overline{R}\right)}{\frac{1}{T}\sum_{t=1}^{T}\left(R_t - \overline{R}\right)^2}, \quad \tau = 1, 2, \ldots, m < T$$

In order to detect dynamics in a time series, it is very useful to first plot the autocorrelation function (ACF), which plots $\widehat{\rho}_\tau$ on the vertical axis against τ on the horizontal axis.

The statistical significance of a set of autocorrelations can be formally tested using the Ljung-Box statistic. It tests the null hypothesis that the autocorrelation for lags 1 through m are all jointly zero via

$$LB(m) = T(T+2)\sum_{\tau=1}^{m}\frac{\widehat{\rho}_\tau^2}{T-\tau} \sim \chi_m^2$$

where χ_m^2 denotes the chi-squared distribution with m degrees of freedom.

The critical value of χ_m^2 corresponding to the probability p can be found for example by using the CHIINV function in Excel. If $p = 0.95$ and $m = 20$, then the formula CHIINV(0.95,20) in Excel returns the value 10.85. If the test statistic $LB(20)$ computed using the first 20 autocorrelations is larger than 10.85 then we reject the hypothesis that the first 20 autocorrelations are zero at the 5% significance level.

Clearly, the maximum number of lags, m, must be chosen in order to implement the test. Often the application at hand will give some guidance. For example if we are looking to detect intramonth dynamics in a daily return, we use $m = 21$ corresponding to 21 trading days in a month. When no such guidance is available, setting $m = \ln(T)$ has been found to work well in simulation studies.

4.2 Autoregressive (AR) Models

Once a pattern has been found in the autocorrelations then we want to build forecasting models that can match the pattern in the autocorrelation function.

The simplest and most used model for this purpose is the autoregressive model of order 1, AR(1), which is defined as

$$R_t = \phi_0 + \phi_1 R_{t-1} + \varepsilon_t$$

where $E[\varepsilon_t] = 0$, $Var[\varepsilon_t] = \sigma_\varepsilon^2$ and where we assume that $R_{t-\tau}$ and ε_t are independent for all $\tau > 0$. Under these assumptions the conditional mean forecast for one period ahead is

$$E(R_{t+1}|R_t) = E\left(\phi_0 + \phi_1 R_t + \varepsilon_{t+1}|R_t\right) = \phi_0 + \phi_1 R_t$$

By writing the AR(1) model for $R_{t+\tau}$ and repeatedly substituting past values we get

$$\begin{aligned} R_{t+\tau} &= \phi_0 + \phi_1 R_{t+\tau-1} + \varepsilon_{t+\tau} \\ &= \phi_0 + \phi_1^2 R_{t+\tau-2} + \phi_1 \varepsilon_{t+\tau-1} + \varepsilon_{t+\tau} \\ &\quad \ldots \\ &= \phi_0 + \phi_1^\tau R_t + \phi_1^{\tau-1} \varepsilon_{t+1} + \cdots + \phi_1 \varepsilon_{t+\tau-1} + \varepsilon_{t+\tau} \end{aligned}$$

The multistep forecast in the AR(1) model is therefore

$$E(R_{t+\tau}|R_t) = \phi_0 + \phi_1^\tau R_t$$

If $|\phi_1| < 1$ then the (unconditional) mean of the model can be denoted by

$$E(R_t) = E(R_{t-1}) = \mu$$

which in the AR(1) model implies

$$\begin{aligned} E(R_t) &= \phi_0 + \phi_1 E(R_{t-1}) + E(\varepsilon_t) \\ \mu &= \phi_0 + \phi_1 \mu, \text{ and so} \\ E(R_t) &= \mu = \frac{\phi_0}{1 - \phi_1} \end{aligned}$$

The unconditional variance is similarly

$$\begin{aligned} Var(R_t) &= \phi_1^2 Var(R_{t-1}) + Var(\varepsilon_t), \text{ so that} \\ Var(R_t) &= \frac{\sigma_\varepsilon^2}{1 - \phi_1^2} \end{aligned}$$

because $Var(R_t) = Var(R_{t-1})$ when $|\phi_1| < 1$.

Just as time series data can be characterized by the ACF then so can linear time series models. To derive the ACF for the AR(1) model assume without loss of generality that $\mu = 0$. Then

$$
\begin{aligned}
R_t &= \phi_1 R_{t-1} + \varepsilon_t, \text{ and} \\
R_t R_{t-\tau} &= \phi_1 R_{t-1} R_{t-\tau} + \varepsilon_t R_{t-\tau}, \text{ and so} \\
E(R_t R_{t-\tau}) &= \phi_1 E(R_{t-1} R_{t-\tau}), \text{ which implies} \\
\rho_\tau &= \phi_1 \rho_{\tau-1}, \text{ so that} \\
\rho_\tau &= \phi_1^\tau \rho_0 = \phi_1^\tau
\end{aligned}
$$

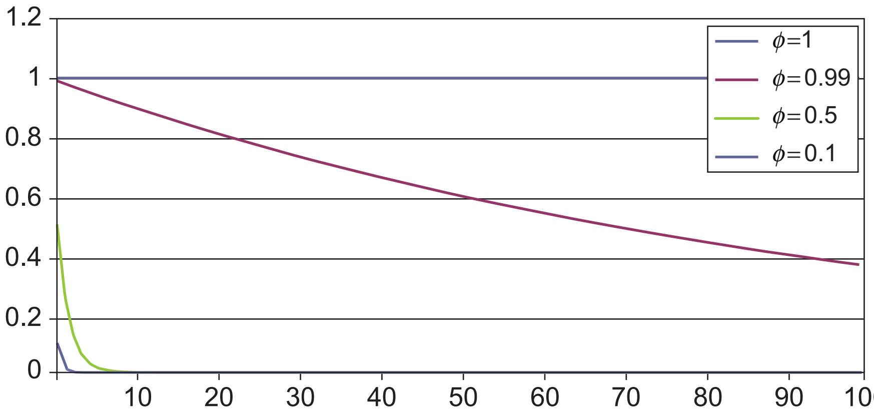
This provides the ACF of the AR(1) model. Notice the similarity between the ACF and the multistep forecast earlier.

The lag order τ appears in the exponent of ϕ_1 and we therefore say that the ACF of an AR(1) model decays exponentially to zero as τ increases. The case when ϕ_1 is close to 1 but not quite 1 is important in financial economics. We refer to this as a highly persistent series.

Figure 3.2 shows examples of the ACF in AR(1) models with four different (positive) values of ϕ_1. When $\phi_1 < 1$ then the ACF decays to zero exponentially. Clearly the decay is much slower when $\phi_1 = 0.99$ than when it is 0.5 or 0.1. When $\phi_1 = 1$ then the ACF is flat at 1. This is the case of a random walk, which we will study further later.

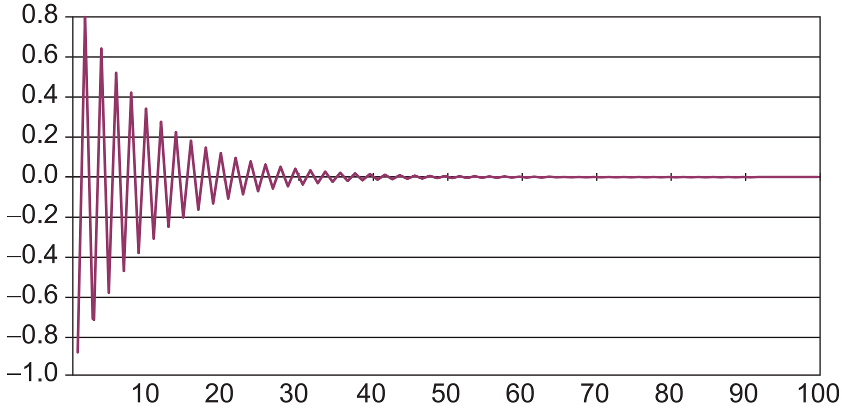
Figure 3.3 shows the ACF of an AR(1) when $\phi_1 = -0.9$. Notice the drastically different ACF pattern compared with Figure 3.2. When $\phi_1 < 0$ then the ACF oscillates around zero but it still decays to zero as the lag order increases. The ACFs in Figure 3.2 are much more common in financial risk management than are the ACFs in Figure 3.3.

Figure 3.2 Autocorrelation functions for AR(1) models with positive ϕ_1.



Notes: We plot the autocorrelation function for four AR(1) processes with different values of the autoregressive parameter ϕ_1. When $\phi_1 < 1$ then the ACF decays to 0 at an exponential rate.

Figure 3.3 Autocorrelation functions for an AR(1) model with $\phi = -0.9$.



Notes: We plot the autocorrelation function for an AR(1) model with $\phi = -0.9$ against the lag order.

The simplest extension to the AR(1) model is the AR(2) defined as

$$R_t = \phi_0 + \phi_1 R_{t-1} + \phi_2 R_{t-2} + \varepsilon_t$$

The autocorrelation function of the AR(2) is

$$\rho_\tau = \phi_1 \rho_{\tau-1} + \phi_2 \rho_{\tau-2}, \quad \text{for } \tau > 1$$

for example

$$E(R_t R_{t-3}) = \phi_1 E(R_{t-1} R_{t-3}) + \phi_2 E(R_{t-2} R_{t-3})$$

so that

$$\rho_3 = \phi_1 \rho_2 + \phi_2 \rho_1$$

In order to derive the first-order autocorrelation note first that the ACF is symmetric around $\tau = 0$ meaning that

$$Corr(R_t, R_{t-\tau}) = Corr\left(R_{t,} R_{t+\tau}\right) \quad \text{for all } \tau$$

We therefore get that

$$E(R_t R_{t-1}) = \phi_1 E(R_{t-1} R_{t-1}) + \phi_2 E(R_{t-2} R_{t-1})$$

implies

$$\rho_1 = \phi_1 + \phi_2 \rho_1$$

so that

$$\rho_1 = \frac{\phi_1}{1 - \phi_2}$$

The general AR(p) model is defined by

$$R_t = \phi_0 + \phi_1 R_{t-1} + \cdots + \phi_p R_{t-p} + \varepsilon_t$$

The one-step ahead forecast in the AR(p) model is simply

$$E_t(R_{t+1}) \equiv E\big(R_{t+1}|R_t, R_{t-1}, \ldots\big) = \phi_0 + \phi_1 R_t + \cdots + \phi_p R_{t+1-p}$$

The τ day ahead forecast can be built using

$$E_t\left(R_{t+\tau}\right) = \phi_0 + \sum_{i=1}^{p} \phi_i E_t(R_{t+\tau-i})$$

which is sometimes called the chain-rule of forecasting. Note that when $\tau < i$ then

$$E_t(R_{t+\tau-i}) = R_{t+\tau-i}$$

because $R_{t+\tau-i}$ is known at the time the forecast is made when $\tau < i$.

The partial autocorrelation function (PACF) gives the marginal contribution of an additional lagged term in AR models of increasing order. First estimate a series of AR models of increasing order:

$$\begin{aligned} R_t &= \phi_{0,1} + \phi_{1,1} R_{t-1} + \varepsilon_{1t} \\ R_t &= \phi_{0,2} + \phi_{1,2} R_{t-1} + \phi_{2,2} R_{t-2} + \varepsilon_{2t} \\ R_t &= \phi_{0,3} + \phi_{1,3} R_{t-1} + \phi_{2,3} R_{t-2} + \phi_{3,3} R_{t-3} + \varepsilon_{3t} \\ &\vdots \; \vdots \end{aligned}$$

The PACF is now defined as the collection of the largest order coefficients

$$\left\{\phi_{1,1}, \phi_{2,2}, \phi_{3,3}, \ldots\right\}$$

which can be plotted against the lag order just as we did for the ACF.

The optimal lag order p in the AR(p) can be chosen as the largest p such that $\phi_{p,p}$ is significant in the PACF. For example, an AR(3) will have a significant $\phi_{3,3}$ but it will have a $\phi_{4,4}$ close to zero.

Note that in AR models the ACF decays exponentially whereas the PACF decays abruptly. This is why the PACF is useful for AR model order selection.

The AR(p) models can be easily estimated using simple OLS regression on observations $p+1$ through T. A useful diagnostic test of the model is to plot the ACF of residuals from the model and perform a Ljung-Box test on the residuals using $m-p$ degrees of freedom.

4.3 Moving Average (MA) Models

In AR models the ACF dies off exponentially, however, certain dynamic features such as bid-ask bounces or measurement errors die off abruptly and require a different type of model. Consider the MA(1) model in which

$$R_t = \theta_0 + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

where ε_t and ε_{t-1} are independent of each other and where $E[\varepsilon_t] = 0$. Note that

$$E[R_t] = \theta_0$$

and

$$Var(R_t) = (1+\theta_1^2)\sigma_\varepsilon^2$$

In order to derive the ACF of the MA(1) assume without loss of generality that $\theta_0 = 0$. We then have

$$\begin{aligned} R_t &= \varepsilon_t + \theta_1 \varepsilon_{t-1} \text{ which implies} \\ R_{t-\tau} R_t &= R_{t-\tau}\varepsilon_t + \theta_1 R_{t-\tau}\varepsilon_{t-1}, \text{ so that} \\ E(R_{t-1}R_t) &= \theta_1 \sigma_\varepsilon^2, \text{ and} \\ E(R_{t-\tau}R_t) &= 0, \quad \text{for } \tau > 1 \end{aligned}$$

Using the variance expression from before, we get the ACF

$$\begin{aligned} \rho_1 &= \frac{\theta_1}{1+\theta_1^2}, \text{ and} \\ \rho_\tau &= 0, \quad \text{for } \tau > 1 \end{aligned}$$

Note that the autocorrelations for the MA(1) are zero for $\tau > 1$.

Unlike AR models, the MA(1) model must be estimated by numerical optimization of the likelihood function. We proceed as follows. First, set the unobserved $\varepsilon_0 = 0$, which is its expected value. Second, set parameter starting values (initial guesses) for θ_0, θ_1, and σ_ε^2. We can use the average of R_t for θ_0, use 0 for θ_1, and use the sample variance of R_t for σ_ε^2. Now we can compute the time series of residuals via

$$\varepsilon_t = R_t - \theta_0 - \theta_1 \varepsilon_{t-1}, \quad \text{with } \varepsilon_0 = 0$$

We are now ready to estimate the parameters by maximizing the likelihood function that we must first define. Let us first assume that ε_t is normally distributed, then

$$f(\varepsilon_t) = \frac{1}{\left(2\pi\sigma_\varepsilon^2\right)^{1/2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_\varepsilon^2}\right)$$

To construct the likelihood function note that as the ε_ts are independent over time we have

$$f(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T) = f(\varepsilon_1) f(\varepsilon_2) \ldots f(\varepsilon_T)$$

and we therefore can write the joint distribution of the sample as

$$f(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T) = \prod_{t=1}^{T} \frac{1}{\left(2\pi\sigma_\varepsilon^2\right)^{1/2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_\varepsilon^2}\right)$$

The maximum likelihood estimation method chooses parameters to maximize the probability of the estimated model (in this case MA(1)) having generated the observed data set (in this case the set of R_ts).

In the MA(1) model we must perform an iterative search (using for example Solver in Excel) over the parameters $\theta_0, \theta_1, \sigma_\varepsilon^2$:

$$L\left(R_1, \ldots, R_T | \theta_0, \theta_1, \sigma_\varepsilon^2\right) = \prod_{t=1}^{T} \frac{1}{\left(2\pi\sigma_\varepsilon^2\right)^{1/2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_\varepsilon^2}\right)$$

$$\text{where } \varepsilon_t = R_t - \theta_0 - \theta_1 \varepsilon_{t-1}, \text{ with } \varepsilon_0 = 0$$

Once the parameters have been estimated we can use the model for forecasting. In the MA(1) model the conditional mean forecast is

$$E(R_{t+1} | R_t, R_{t-1}, \ldots) = \theta_0 + \theta_1 \varepsilon_t$$
$$E(R_{t+\tau} | R_t, R_{t-1}, \ldots) = \theta_0, \quad \text{for } \tau > 1$$

The general MA(q) model is defined by

$$R_t = \theta_0 + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

It has an ACF that is nonzero for the first q lags and then zero for lags larger than q.

Note that MA models are easily identified using the ACF. If the ACF of a data series dies off to zero abruptly after the first four (nonzero) lags then an MA(4) is likely to provide a good fit of the data.

4.4 Combining AR and MA into ARMA Models

Parameter parsimony is key in forecasting, and combining AR and MA models into ARMA models often enables us to model dynamics with fewer parameters.

Consider the ARMA(1,1) model, which includes one lag of R_t and one lag of ε_t:

$$R_t = \phi_0 + \phi_1 R_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

As in the AR(1), the mean of the ARMA(1,1) time series is given from

$$E[R_t] = \phi_0 + \phi_1 E[R_{t-1}] = \phi_0 + \phi_1 E[R_t]$$

which implies that

$$E(R_t) = \frac{\phi_0}{1-\phi_1}$$

when $\left|\phi_1\right| < 1$. In this case R_t will tend to fluctuate around the mean, $\phi_0/\left(1-\phi_1\right)$, over time. We say that R_t is mean-reverting in this case.

Using the fact that $E[R_t\varepsilon_t] = \sigma_\varepsilon^2$ we can get the variance from

$$Var[R_t] = \phi_1^2 Var[R_t] + \theta_1^2\sigma_\varepsilon^2 + \sigma_\varepsilon^2 + 2\phi_1\theta_1\sigma_\varepsilon^2$$

which implies that

$$Var(R_t) = \frac{(1+2\phi_1\theta_1+\theta_1^2)\sigma_\varepsilon^2}{1-\phi_1^2}$$

The first-order autocorrelation is given from

$$E[R_tR_{t-1}] = \phi_1 E[R_{t-1}R_{t-1}] + \theta_1 E[\varepsilon_{t-1}R_{t-1}] + E[\varepsilon_t R_{t-1}]$$

in which we assume again that $\phi_0 = 0$. This implies that

$$\rho_1 Var(R_t) = \phi_1 Var(R_t) + \theta_1\sigma_\varepsilon^2$$

so that

$$\rho_1 = \phi_1 + \frac{\theta_1\sigma_\varepsilon^2}{Var(R_t)}$$

For higher order autocorrelations the MA term has no effect and we get the same structure as in the AR(1) model

$$\rho_\tau = \phi_1\rho_{\tau-1}, \quad \text{for } \tau > 1$$

The general ARMA(p,q) model is

$$R_t = \phi_0 + \sum_{i=1}^{p}\phi_i R_{t-i} + \sum_{i=1}^{q}\theta_i\varepsilon_{t-i} + \varepsilon_t$$

Because of the MA term, ARMA models just as MA models must be estimated using maximum likelihood estimation (MLE). Diagnostics on the residuals can be done via Ljung-Box tests with degrees of freedom equal to $m-q-p$.

4.5 Random Walks, Units Roots, and ARIMA Models

The random walk model is a key benchmark in financial forecasting. It is often used to model speculative prices in logs. Let S_t be the closing price of an asset and let $s_t = \ln(S_t)$ so that log returns are immediately defined by $R_t \equiv \ln(S_t) - \ln(S_{t-1}) = s_t - s_{t-1}$.

The random walk (or martingale) model for log prices is now defined by

$$s_t = s_{t-1} + \varepsilon_t$$

By iteratively substituting in lagged log prices we can write

$$\begin{aligned} s_t &= s_{t-2} + \varepsilon_{t-1} + \varepsilon_t \\ &\dots \\ s_t &= s_{t-\tau} + \varepsilon_{t-\tau+1} + \varepsilon_{t-\tau+2} + \cdots + \varepsilon_t \end{aligned}$$

Because past $\varepsilon_{t-\tau}$ residual (or shocks) matter equally and fully for s_t regardless of τ we say that past shocks have permanent effects in the random walk model.

In the random walk model, the conditional mean and variance forecasts for the log price are

$$\begin{aligned} E_t(s_{t+\tau}) &= s_t \\ Var_t(s_{t+\tau}) &= \tau\sigma_\varepsilon^2 \end{aligned}$$

Note that the forecast for s at any horizon is just today's value, s_t. We therefore sometimes say that the random walk model implies that the series is not predictable. Note also that the conditional variance of the future value is a linear function of the forecast horizon, τ.

Equity returns typically have a small positive mean corresponding to a small drift in the log price. This motivates the random walk with drift model

$$s_t = \mu + s_{t-1} + \varepsilon_t$$

Substituting in lagged prices back to time 0, we have

$$s_t = t\mu + s_0 + \varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_1$$

Notice that in this model the constant drift μ in returns corresponds to a coefficient on time, t, in the log price model. We call this a deterministic time trend and we refer to the sum of the εs as a stochastic trend.

A time series, s_t, follows an ARIMA$(p, 1, q)$ model if the first differences, $s_t - s_{t-1}$, follow a mean-reverting ARMA(p, q) model. In this case we say that s_t has a unit root. The random walk model has a unit root as well because in that model

$$s_t - s_{t-1} = \varepsilon_t$$

which is a trivial ARMA(0,0) model.

4.6 Pitfall 1: Spurious Mean-Reversion

Consider the AR(1) model again:

$$
\begin{aligned}
s_t &= \phi_1 s_{t-1} + \varepsilon_t \Leftrightarrow \\
s_t - s_{t-1} &= (\phi_1 - 1)s_{t-1} + \varepsilon_t
\end{aligned}
$$

Note that when $\phi_1 = 1$ then the AR(1) model has a unit root and becomes the random walk model. The OLS estimator contains an important small sample bias in dynamic models. For example, in an AR(1) model when the true ϕ_1 coefficient is close or equal to 1, the finite sample OLS estimate will be biased downward. This is known as the Hurwitz bias or the Dickey-Fuller bias. This bias is important to keep in mind.

If ϕ_1 is estimated in a small sample of asset prices to be 0.85 then it implies that the underlying asset price is predictable and market timing thus feasible. However, the true value may in fact be 1, which means that the price is a random walk and so unpredictable.

The aim of technical trading analysis is to find dynamic patterns in asset prices. Econometricians are very skeptical about this type of analysis exactly because it attempts to find dynamic patterns in prices and not returns. Asset prices are likely to have a ϕ_1 very close to 1, which in turn is likely to be estimated to be somewhat lower than 1, which in turn suggests predictability. Asset returns have a ϕ_1 close to zero and the estimate of an AR(1) on returns does not suffer from bias. Looking for dynamic patterns in asset returns is much less likely to produce false evidence of predictability than is looking for dynamic patterns in asset returns. Risk managers ought to err on the side of prudence and thus consider dynamic models of asset returns and not asset prices.

4.7 Testing for Unit Roots

Asset prices often have a ϕ_1 very close to 1. But we are very interested in knowing whether $\phi_1 = 0.99$ or 1 because the two values have very different implications for longer term forecasting as indicated by Figure 3.2. $\phi_1 = 0.99$ implies that the asset price is predictable so that market timing is possible whereas $\phi_1 = 1$ implies it is not. Consider again the AR(1) model with and without a constant term:

$$
\begin{aligned}
s_t &= \phi_0 + \phi_1 s_{t-1} + \varepsilon_t \\
s_t &= \phi_1 s_{t-1} + \varepsilon_t
\end{aligned}
$$

Unit root tests (also known as Dickey-Fuller tests) have been developed to assess the null hypothesis

$$H_0 : \phi_1 = 1$$

against the alternative hypothesis that

$$H_A : \phi_1 < 1$$

This looks like a standard t-test in a regression but it is crucial that when the null hypothesis H_0 is true, so that $\phi_1 = 1$, the unit root test does not have the usual normal distribution even when T is large. If you estimate ϕ_1 using OLS and test that $\phi_1 = 1$ using the usual t-test with critical values from the normal distribution then you are likely to reject the null hypothesis much more often than you should. This means that you are likely to spuriously find evidence of mean-reversion, that is, predictability.

5 Multivariate Time Series Models

Multivariate time series analysis is relevant for risk management because we often consider risk models with multiple related risk factors or models with many assets. This section will briefly introduce the following important topics: time series regressions, spurious relationships, cointegration, cross correlations, vector autoregressions, and spurious causality.

5.1 Time Series Regression

The relationship between two (or more) time series can be assessed applying the usual regression analysis. But in time series analysis the regression errors must be scrutinized carefully.

Consider a simple bivariate regression of two highly persistent series, for example, the spot and futures price of an asset

$$s_{1t} = a + bs_{2t} + e_t$$

The first step in diagnosing such a time series regression model is to plot the ACF of the regression errors, e_t.

If ACF dies off only very slowly (the Hurwitz bias will make the ACF look like it dies off faster to zero than it really does) then it is good practice to first-difference each series and run the regression

$$(s_{1t} - s_{1t-1}) = a + b\,(s_{2t} - s_{2t-1}) + e_t$$

Now the ACF can be used on the residuals of the new regression and the ACF can be checked for dynamics. The AR, MA, or ARMA models can be used to model any dynamics in e_t. After modeling and estimating the parameters in the residual time series, e_t, the entire regression model including a and b can be reestimated using MLE.

5.2 Pitfall 2: Spurious Regression

Checking the ACF of the error term in time series regressions is particularly important due to the so-called spurious regression phenomenon: Two completely unrelated times series—each with a unit root—are likely to appear related in a regression that has a significant b coefficient.

Specifically, let s_{1t} and s_{2t} be two independent random walks

$$s_{1t} = s_{1t-1} + \varepsilon_{1t}$$
$$s_{2t} = s_{2t-1} + \varepsilon_{2t}$$

where ε_{1t} and ε_{2t} are independent of each other and independent over time. Clearly the true value of b is zero in the time series regression

$$s_{1t} = a + bs_{2t} + e_t$$

However, in practice, standard t-tests using the estimated b coefficient will tend to conclude that b is nonzero when in truth it is zero. This problem is known as spurious regression.

Fortunately, as noted earlier, the ACF comes to the rescue for detecting spurious regression. If the relationship between s_{1t} and s_{2t} is spurious then the error term, e_t, will have a highly persistent ACF and the regression in first differences

$$(s_{1t} - s_{1t-1}) = a + b\,(s_{2t} - s_{2t-1}) + e_t$$

will not show a significant estimate of b. Note that Pitfall 1, earlier, was related to modeling univariate asset prices time series in levels rather than in first differences. Pitfall 2 is in the same vein: Time series regression on highly persistent asset prices is likely to lead to false evidence of a relationship, that is, a spurious relationship. Regression on returns is much more likely to lead to sensible conclusions about dependence across assets.

5.3 Cointegration

Relationships between variables with unit roots are of course not always spurious. A variable with a unit root, for example a random walk, is also called integrated, and if two variables that are both integrated have a linear combination with no unit root then we say they are cointegrated.

Examples of cointegrated variables could be long-run consumption and production in an economy, or the spot and the futures price of an asset that are related via a no-arbitrage condition. Similarly, consider the pairs trading strategy that consists of finding two stocks whose prices tend to move together. If prices diverge then we buy the temporarily cheap stock and short sell the temporarily expensive stock and wait for the typical relationship between the prices to return. Such a strategy hinges on the stock prices being cointegrated.

Consider a simple bivariate model where

$$s_{1t} = \phi_0 + s_{1,t-1} + \varepsilon_{1t}$$
$$s_{2t} = bs_{1t} + \varepsilon_{2t}$$

Note that s_{1t} has a unit root and that the level of s_{1t} and s_{2t} are related via b. Assume that ε_{1t} and ε_{2t} are independent of each other and independent over time.

The cointegration model can be used to preserve the relationship between the variables in the long-term forecasts

$$E\left(s_{1,t+\tau}|s_{1t},s_{2t}\right) = \phi_0\tau + s_{1t}$$
$$E\left(s_{2,t+\tau}|s_{1t},s_{2t}\right) = b\phi_0\tau + bs_{1t}$$

The concept of cointegration was developed by Rob Engle and Clive Granger. They together received the Nobel Prize in Economics in 2003 for this and many other contributions to financial time series analysis.

5.4 Cross-Correlations

Consider again two financial time series, $R_{1,t}$ and $R_{2,t}$. They can be dependent in three possible ways: $R_{1,t}$ can lead $R_{2,t}$ (e.g., $Corr\left(R_{1,t}, R_{2,t+1}\right) \neq 0$), $R_{1,t}$ can lag $R_{2,t}$ (e.g., $Corr\left(R_{1,t+1}, R_{2,t}\right) \neq 0$), and they can be contemporaneously related (e.g., $Corr\left(R_{1,t}, R_{2,t}\right) \neq 0$). We need a tool to detect all these possible dynamic relationships.

The sample cross-correlation matrices are the multivariate analogues of the ACF function and provide the tool we need. For a bivariate time series, the cross-covariance matrix for lag τ is

$$\Gamma_\tau = \begin{bmatrix} Cov\left(R_{1,t}, R_{1,t-\tau}\right) & Cov\left(R_{1,t}, R_{2,t-\tau}\right) \\ Cov\left(R_{2,t}, R_{1,t-\tau}\right) & Cov\left(R_{2,t}, R_{2,t-\tau}\right) \end{bmatrix}, \quad \tau \geq 0$$

Note that the two diagonal terms are the autocovariance function of $R_{1,t}$, and $R_{2,t}$, respectively.

In the general case of a k-dimensional time series, we have

$$\Gamma_\tau = E\left\{(R_t - E[R_t])(R_{t-\tau} - E[R_t])'\right\}, \quad \tau \geq 0$$

where R_t is now a k by 1 vector of variables.

Detecting lead and lag effects is important, for example when relating an illiquid stock to a liquid market factor. The illiquidity of the stock implies price observations that are often stale, which in turn will have a spuriously low correlation with the liquid market factor. The stale equity price will be correlated with the lagged market factor and this lagged relationship can be used to compute a liquidity-corrected measure of the dependence between the stock and the market.

5.5 Vector Autoregressions (VAR)

The vector autoregression model (VAR), which is not to be confused with Value-at-Risk (VaR), is arguably the simplest and most often used multivariate time series model for forecasting. Consider a first-order VAR, call it VAR(1)

$$R_t = \phi_0 + \Phi R_{t-1} + \varepsilon_t, \quad Var(\varepsilon_t) = \Sigma$$

where R_t is again a k by 1 vector of variables.

The bivariate case is simply

$$R_{1,t} = \phi_{0,1} + \Phi_{11}R_{1,t-1} + \Phi_{12}R_{2,t-1} + \varepsilon_{1,t}$$
$$R_{2,t} = \phi_{0,1} + \Phi_{21}R_{1,t-1} + \Phi_{22}R_{2,t-1} + \varepsilon_{2,t}$$
$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

Note that in the VAR, $R_{1,t}$ and $R_{2,t}$ are contemporaneously related via their covariance $\sigma_{12} = \sigma_{21}$. But just as in the AR model, the VAR only depends on lagged variables so that it is immediately useful in forecasting.

If the variables included on the right-hand-side of each equation in the VAR are the same (as they are above) then the VAR is called unrestricted and OLS can be used equation-by-equation to estimate the parameters.

5.6 Pitfall 3: Spurious Causality

We may sometimes be interested to see if the lagged value of $R_{2,t}$, namely $R_{2,t-1}$, is causal for the current value of $R_{1,t}$, in which case it can be used in forecasting. To this end a simple regression of the form

$$R_{1,t} = a + bR_{2,t-1} + e_t$$

could be used. Note that it is the lagged value $R_{2,t-1}$ that appears on the right-hand side. Unfortunately, such a regression may easily lead to false conclusions if $R_{1,t}$ is persistent and so depends on its own past value, which is not included on the right-hand side of the regression.

In order to truly assess if $R_{2,t-1}$ causes $R_{1,t}$ (or vice versa), we should ask the question: Is past $R_{2,t}$ useful for forecasting current $R_{1,t}$ once the past $R_{1,t}$ has been accounted for? This question can be answered by running a VAR model:

$$R_{1,t} = \phi_{0,1} + \Phi_{11}R_{1,t-1} + \Phi_{12}R_{2,t-1} + \varepsilon_{1,t}$$
$$R_{2,t} = \phi_{0,2} + \Phi_{21}R_{1,t-1} + \Phi_{22}R_{2,t-1} + \varepsilon_{2,t}$$

Now we can define Granger causality (as opposed to spurious causality) as follows:

- $R_{2,t}$ is said to Granger cause $R_{1,t}$ if $\Phi_{12} \neq 0$
- $R_{1,t}$ is said to Granger cause $R_{2,t}$ if $\Phi_{21} \neq 0$

In some cases several lags of $R_{1,t}$ may be needed on the right-hand side of the equation for $R_{1,t}$ and similarly we may need more lags of $R_{2,t}$ in the equation for $R_{2,t}$.

6 Summary

The financial asset prices and portfolio values typically studied by risk managers can be viewed as examples of very persistent time series. An important goal of this chapter

is therefore to ensure that the risk manager avoids some common pitfalls that arise because of the persistence in prices. The three most important issues are

- Spurious detection of mean-reversion; that is, erroneously finding that a variable is mean-reverting when it is truly a random walk
- Spurious regression; that is, erroneously finding that a variable x is significant when regressing y on x
- Spurious detection of causality; that is, erroneously finding that the current value of x causes (helps determine) future values of y when in reality it cannot

Several more advanced topics have been left out of the chapter including long memory models and models of seasonality. Long memory models give more flexibility in modeling the autocorrelation function (ACF) than do the traditional ARIMA and ARMA models studied in this chapter. In particular long-memory models allow for the ACF to go to zero more slowly than the AR(1) model, which decays to zero at an exponential decay as we saw earlier. Seasonal models are useful, for example, for the analysis of agricultural commodity prices where seasonal patterns in supply cause seasonal patterns in prices, in expected returns, and in volatility. These topics can be studied using the resources suggested next.

Further Resources

For a basic introduction to financial data analysis, see Koop (2006) and for an introduction to probability theory see Paollela (2006). Wooldridge (2002) and Stock and Watson (2010) provide a broad introduction to econometrics. Anscombe (1973) contains the data in Table 3.1 and Figure 3.1.

The univariate and multivariate time series material in this chapter is based on Chapters 2 and 8 in Tsay (2002), which should be consulted for various extensions including seasonality and long memory. See also Taylor (2005) for an excellent treatment of financial time series analysis focusing on volatility modeling.

Diebold (2004) gives a thorough introduction to forecasting in economics. Granger and Newbold (1986) is the classic text for the more advanced reader. Christoffersen and Diebold (1998) analyze long-horizon forecasting in cointegrated systems.

The classic references on the key time series topics in this chapter are Hurwitz (1950) on the bias in the AR(1) coefficient, Granger and Newbold (1974) on spurious regression in economics, Engle and Granger (1987) on cointegration, Granger (1969) on Granger causality, and Dickey and Fuller (1979) on unit root testing. Hamilton (1994) provides an authoritative treatment of economic time series analysis.

Tables with critical values for unit root tests can be found in MacKinnon (1996, 2010). See also Chapter 14 in Davidson and MacKinnon (2004).

References

Anscombe, F.J., 1973. Graphs in statistical analysis. Am. Stat. 27, 17–21.

Christoffersen, P., Diebold, F., 1998. Cointegration and long horizon forecasting. J. Bus. Econ. Stat. 16, 450–458.

Davidson, R., MacKinnon, J.G., 2004. Econometric Theory and Methods. Oxford University Press, New York, NY.
Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. J. Am. Stat. Assoc. 74, 427–431.
Diebold, F.X., 2004. Elements of Forecasting, third ed. Thomson South-Western, Cincinnati, Ohio.
Engle, R.F., Granger, C.W.J., 1987. Co-integration and error correction: Representation, estimation and testing. Econometrica 55, 251–276.
Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–438.
Granger, C.W.J., Newbold, P., 1974. Spurious regressions in econometrics. J. Econom. 2, 111–120.
Granger, C.W.J., Newbold, P., 1986. Forecasting Economic Time Series, second ed. Academic Press, Orlando, FL.
Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton, NJ.
Hurwitz, L., 1950. Least squares bias in time series. In: Koopmans, T.C. (Ed.), Statistical Inference in Econometric Models. Wiley, New York, NY.
Koop, G., 2006. Analysis of Financial Data. Wiley, Chichester, West Sussex, England.
MacKinnon, J.G., 1996. Numerical distribution functions for unit root and cointegration tests. J. Appl. Econom. 11, 601–618.
MacKinnon, J.G., 2010. Critical Values for Cointegration Tests, Queen's Economics Department. Working Paper no 1227. http://ideas.repec.org/p/qed/wpaper/1227.html.
Paollela, M., 2006. Fundamental Probability. Wiley, Chichester, West Sussex, England.
Stock, J., Watson, M., 2010. Introduction to Econometrics, second ed. Pearson Addison Wesley.
Taylor, S.J., 2005. Asset Price Dynamics, Volatility and Prediction. Princeton University Press, Princeton, NJ.
Tsay, R., 2002. Analysis of Financial Time Series. Wiley Interscience, Hoboken, NJ.
Wooldridge, J., 2002. Introductory Econometrics: A Modern Approach. Second Edition. South-Western College Publishing, Mason, Ohio.

Empirical Exercises

Open the Chapter3Data.xlsx file from the web site.

1. Using the data in the worksheet named Question 3.1 reproduce the moments and regression coefficients at the bottom of Table 3.1.
2. Reproduce Figure 3.1.
3. Reproduce Figure 3.2.
4. Using the data sets in the worksheet named Question 3.4, estimate an AR(1) model on each of the 100 columns of data. (*Excel hint*: Use the LINEST function.) Plot the histogram of the 100 ϕ_1 estimates you have obtained. The true value of ϕ_1 is one in all the columns. What does the histogram tell you?
5. Using the data set in the worksheet named Question 3.4, estimate an MA(1) model using maximum likelihood. Use the starting values suggested in the text. Use Solver in Excel to maximize the likelihood function.

Answers to these exercises can be found on the companion site.

For more information see the companion site at
http://www.elsevierdirect.com/companions/9780123744487

4 Volatility Modeling Using Daily Data

1 Chapter Overview

Part II of the book consists of three chapters. The ultimate goal of this and the following two chapters is to establish a framework for modeling the dynamic distribution of portfolio returns. The methods we develop in Part II can also be used to model each asset in the portfolio separately. In Part III of the book we will consider multivariate models that can link the univariate asset return models together. If the risk manager only cares about risk measurement at the portfolio level then the univariate models in Part II will suffice.

We will proceed with the univariate models in two steps. The first step is to establish a forecasting model for dynamic portfolio variance and to introduce methods for evaluating the performance of these forecasts. The second step is to consider ways to model nonnormal aspects of the portfolio return—that is, aspects that are not captured by the dynamic variance.

The second step, allowing for nonnormal distributions, is covered in Chapter 6. The first step, volatility modeling, is analyzed in this chapter and in Chapter 5. Chapter 5 relies on intraday data to develop daily volatility forecasts. The present chapter focuses on modeling daily volatility when only daily return data are available. We proceed as follows:

1. We briefly describe the simplest variance models available including moving averages and the so-called RiskMetrics variance model.
2. We introduce the GARCH variance model and compare it with the RiskMetrics model.
3. We estimate the GARCH parameters using the quasi-maximum likelihood method.
4. We suggest extensions to the basic model, which improve the model's ability to capture variance persistence and leverage effects. We also consider ways to expand the model, taking into account explanatory variables such as volume effects, day-of-week effects, and implied volatility from options.
5. We discuss various methods for evaluating the volatility forecasting models.

The overall objective of this chapter is to develop a general class of models that can be used by risk managers to forecast daily portfolio volatility using daily return data.

Elements of Financial Risk Management. DOI: 10.1016/B978-0-12-374448-7.00004-X

2 Simple Variance Forecasting

We begin by establishing some notation and by laying out the underlying assumptions for this chapter. In Chapter 1, we defined the daily asset log return, R_{t+1}, using the daily closing price, $S_{t+1,}$ as

$$R_{t+1} \equiv \ln\left(S_{t+1}/S_t\right)$$

We will use the notation R_{t+1} to describe either an individual asset return or the aggregate return on a portfolio. The models in this chapter can be used for both.

We will also apply the finding from Chapter 1 that at short horizons such as daily, we can safely assume that the mean value of R_{t+1} is zero since it is dominated by the standard deviation. Issues arising at longer horizons will be discussed in Chapter 8. Furthermore, we will assume that the innovation to asset return is normally distributed. We hasten to add that the normality assumption is not realistic, and it will be relaxed in Chapter 6. Normality is simply assumed for now, as it allows us to focus on modeling the conditional variance of the distribution.

Given the assumptions made, we can write the daily return as

$$R_{t+1} = \sigma_{t+1} z_{t+1}, \text{with } z_{t+1} \sim \text{i.i.d. } N(0,1)$$

where the abbreviation i.i.d. $N(0,1)$ stands for "independently and identically normally distributed with mean equal to zero and variance equal to 1."

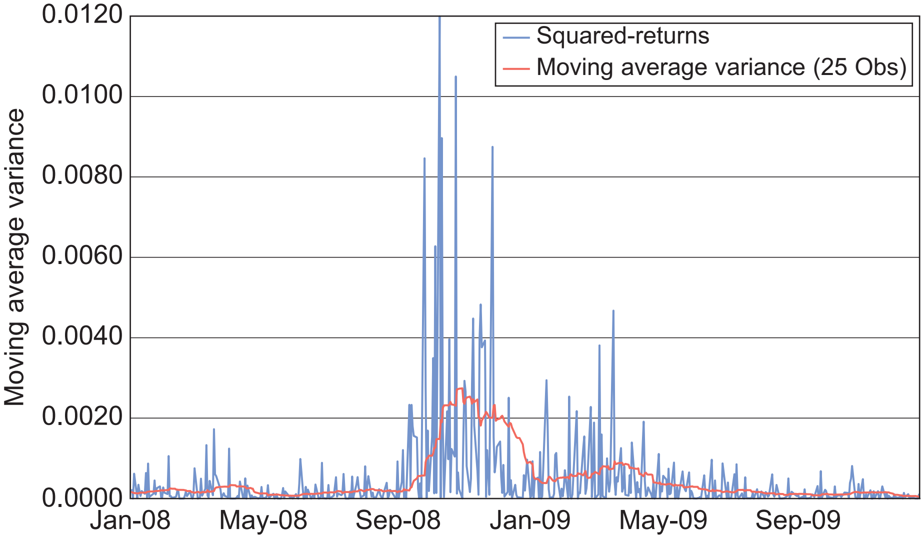
Together these assumptions imply that once we have established a model of the time-varying variance, σ^2_{t+1}, we will know the entire distribution of the asset, and we can therefore easily calculate any desired risk measure. We are well aware from the stylized facts discussed in Chapter 1 that the assumption of conditional normality that is imposed here is not satisfied in actual data on speculative returns. However, as we will see later, for the purpose of variance modeling, we are allowed to assume normality even if it is strictly speaking not a correct assumption. This assumption conveniently allows us to postpone discussions of nonnormal distributions to a later chapter.

The focus of this chapter then is to establish a model for forecasting tomorrow's variance, σ^2_{t+1}. We know from Chapter 1 that variance, as measured by squared returns, exhibits strong autocorrelation, so that if the recent period was one of high variance, then tomorrow is likely to be a high-variance day as well. The easiest way to capture this phenomenon is by letting tomorrow's variance be the simple average of the most recent m observations, as in

$$\sigma^2_{t+1} = \frac{1}{m}\sum_{\tau=1}^{m} R^2_{t+1-\tau} = \sum_{\tau=1}^{m} \frac{1}{m} R^2_{t+1-\tau}$$

Notice that this is a proper forecast in the sense that the forecast for tomorrow's variance is immediately available at the end of today when the daily return is realized. However, the fact that the model puts equal weights (equal to $1/m$) on the past

Figure 4.1 Squared S&P 500 returns with moving average variance estimated on past 25 observations, 2008–2009.



Notes: The daily squared returns are plotted along with a moving average of 25 observations.

m observations yields unwarranted results. An extreme return (either positive or negative) today will bump up variance by $1/m$ times the return squared for exactly m periods after which variance immediately will drop back down. Figure 4.1 illustrates this point for $m = 25$ days. The autocorrelation plot of squared returns in Chapter 1 suggests that a more gradual decline is warranted in the effect of past returns on today's variance. Even if we are content with the box patterns, it is not at all clear how m should be chosen. This is unfortunate as the choice of m is crucial in deciding the patters of σ_{t+1}: A high m will lead to an excessively smoothly evolving σ_{t+1}, and a low m will lead to an excessively jagged pattern of σ_{t+1} over time.

JP Morgan's RiskMetrics system for market risk management considers the following model, where the weights on past squared returns decline exponentially as we move backward in time. The RiskMetrics variance model, or the exponential smoother as it is sometimes called, is written as

$$\sigma_{t+1}^2 = (1 - \lambda) \sum_{\tau=1}^{\infty} \lambda^{\tau-1} R_{t+1-\tau}^2, \quad \text{for } 0 < \lambda < 1$$

Separating from the sum the squared return term for $\tau = 1$, where $\lambda^{\tau-1} = \lambda^0 = 1$, we get

$$\sigma_{t+1}^2 = (1 - \lambda) \sum_{\tau=2}^{\infty} \lambda^{\tau-1} R_{t+1-\tau}^2 + (1 - \lambda) R_t^2$$

Applying the exponential smoothing definition again, we can write today's variance, σ_t^2, as

$$\sigma_t^2 = (1-\lambda)\sum_{\tau=1}^{\infty}\lambda^{\tau-1}R_{t-\tau}^2 = \frac{1}{\lambda}(1-\lambda)\sum_{\tau=2}^{\infty}\lambda^{\tau-1}R_{t+1-\tau}^2$$

so that tomorrow's variance can be written

$$\sigma_{t+1}^2 = \lambda\sigma_t^2 + (1-\lambda)R_t^2$$

The RiskMetrics model's forecast for tomorrow's volatility can thus be seen as a weighted average of today's volatility and today's squared return.

The RiskMetrics model has some clear advantages. First, it tracks variance changes in a way that is broadly consistent with observed returns. Recent returns matter more for tomorrow's variance than distant returns as λ is less than one and therefore the impact of the lagged squared return gets smaller when the lag, τ, gets bigger. Second, the model only contains one unknown parameter, namely, λ. When estimating λ on a large number of assets, RiskMetrics found that the estimates were quite similar across assets, and they therefore simply set $\lambda = 0.94$ for every asset for daily variance forecasting. In this case, no estimation is necessary, which is a huge advantage in large portfolios. Third, relatively little data need to be stored in order to calculate tomorrow's variance. The weight on today's squared returns is $(1-\lambda) = 0.06$, and the weight is exponentially decaying to $(1-\lambda)\lambda^{99} = 0.000131$ on the 100th lag of squared return. After including 100 lags of squared returns, the cumulated weight is $(1-\lambda)\sum_{\tau=1}^{100}\lambda^{\tau-1} = 0.998$, so that 99.8% of the weight has been included. Therefore it is only necessary to store about 100 daily lags of returns in order to calculate tomorrow's variance, σ_{t+1}^2.

Given all these advantages of the RiskMetrics model, why not simply end the discussion on variance forecasting here and move on to distribution modeling? Unfortunately, as we will see shortly, the RiskMetrics model does have certain shortcomings, which will motivate us to consider slightly more elaborate models. For example, it does not allow for a leverage effect, which we considered a stylized fact in Chapter 1, and it also provides counterfactual longer-horizon forecasts.

3 The GARCH Variance Model

We now introduce a set of models that capture important features of returns data and that are flexible enough to accommodate specific aspects of individual assets. The downside of these models is that they require nonlinear parameter estimation, which will be discussed subsequently.

The simplest generalized autoregressive conditional heteroskedasticity (GARCH) model of dynamic variance can be written as

$$\sigma_{t+1}^2 = \omega + \alpha R_t^2 + \beta\sigma_t^2, \quad \text{with } \alpha+\beta < 1$$

Notice that the RiskMetrics model can be viewed as a special case of the simple GARCH model if we force $\alpha = 1-\lambda, \beta = \lambda$, so that $\alpha + \beta = 1$, and further $\omega = 0$. Thus, the two models appear to be quite similar. However, there is an important difference: We can define the unconditional, or long-run average, variance, σ^2, to be

$$\begin{aligned}\sigma^2 &\equiv E[\sigma^2_{t+1}] = \omega + \alpha E[R^2_t] + \beta E[\sigma^2_t] \\ &= \omega + \alpha\sigma^2 + \beta\sigma^2, \text{ so that} \\ \sigma^2 &= \omega/(1-\alpha-\beta)\end{aligned}$$

It is now clear that if $\alpha + \beta = 1$ as is the case in the RiskMetrics model, then the long-run variance is not well defined in that model. Thus, an important quirk of the RiskMetrics model emerges: It ignores the fact that the long-run average variance tends to be relatively stable over time. The GARCH model, in turn, implicitly relies on σ^2. This can be seen by solving for ω in the long-run variance equation and substituting it into the dynamic variance equation. We get

$$\sigma^2_{t+1} = (1-\alpha-\beta)\sigma^2 + \alpha R^2_t + \beta\sigma^2_t = \sigma^2 + \alpha(R^2_t - \sigma^2) + \beta(\sigma^2_t - \sigma^2)$$

Thus, tomorrow's variance is a weighted average of the long-run variance, today's squared return, and today's variance. Put differently, tomorrow's variance is the long-run average variance with something added (subtracted) if today's squared return is above (below) its long-run average, and something added (subtracted) if today's variance is above (below) its long-run average.

Our intuition might tell us that ignoring the long-run variance, as the RiskMetrics model does, is more important for longer-horizon forecasting than for forecasting simply one day ahead. This intuition is correct, as we will now see.

A key advantage of GARCH models for risk management is that the one-day forecast of variance, $\sigma^2_{t+1|t}$, is given directly by the model by σ^2_{t+1}. Consider now forecasting the variance of the daily return k days ahead, using only information available at the end of today. In GARCH, the expected value of future variance at horizon k is

$$\begin{aligned}E_t\left[\sigma^2_{t+k}\right] - \sigma^2 &= \alpha E_t\left[R^2_{t+k-1} - \sigma^2\right] + \beta E_t\left[\sigma^2_{t+k-1} - \sigma^2\right] \\ &= \alpha E_t\left[\sigma^2_{t+k-1} z^2_{t+k-1} - \sigma^2\right] + \beta E_t\left[\sigma^2_{t+k-1} - \sigma^2\right] \\ &= (\alpha+\beta)\left(E_t\left[\sigma^2_{t+k-1}\right] - \sigma^2\right), \text{ so that} \\ E_t\left[\sigma^2_{t+k}\right] - \sigma^2 &= (\alpha+\beta)^{k-1}\left(E_t\left[\sigma^2_{t+1}\right] - \sigma^2\right) = (\alpha+\beta)^{k-1}\left(\sigma^2_{t+1} - \sigma^2\right)\end{aligned}$$

The conditional expectation, $E_t[\bullet]$, refers to taking the expectation using all the information available at the end of day t, which includes the squared return on day t itself.

We will refer to $\alpha + \beta$ as the persistence of the model. A high persistence—that is, an $(\alpha + \beta)$ close to 1—implies that shocks that push variance away from its long-run average will persist for a long time, but eventually the long-horizon forecast will

be the long-run average variance, σ^2. Similar calculations for the RiskMetrics model reveal that

$$E_t\left[\sigma^2_{t+k}\right] = \sigma^2_{t+1}, \forall k$$

as $\alpha + \beta = 1$ and σ^2 is undefined. Thus, persistence in this model is 1, which implies that a shock to variance persists forever: An increase in variance will push up the variance forecast by an identical amount for all future forecast horizons. This is another way of saying that the RiskMetrics model ignores the long-run variance when forecasting. If $\alpha + \beta$ is close to 1 as is typically the case, then the two models might yield similar predictions for short horizons, k, but their longer horizon implications are very different. If today is a high-variance day, then the RiskMetrics model predicts that all future days will be high-variance. The GARCH model more realistically assumes that eventually in the future variance will revert to the average value.

So far we have considered forecasting the variance of daily returns k days ahead. Of more immediate interest is probably the forecast of variance of K-day cumulative returns,

$$R_{t+1:t+K} \equiv \sum_{k=1}^{K} R_{t+k}$$

As we assume that returns have zero autocorrelation, the variance of the cumulative K-day returns is simply

$$\sigma^2_{t+1:t+K} \equiv E_t\left(\sum_{k=1}^{K} R_{t+k}\right)^2 = \sum_{k=1}^{K} E_t\left[\sigma^2_{t+k}\right]$$

So in the RiskMetrics model, we get

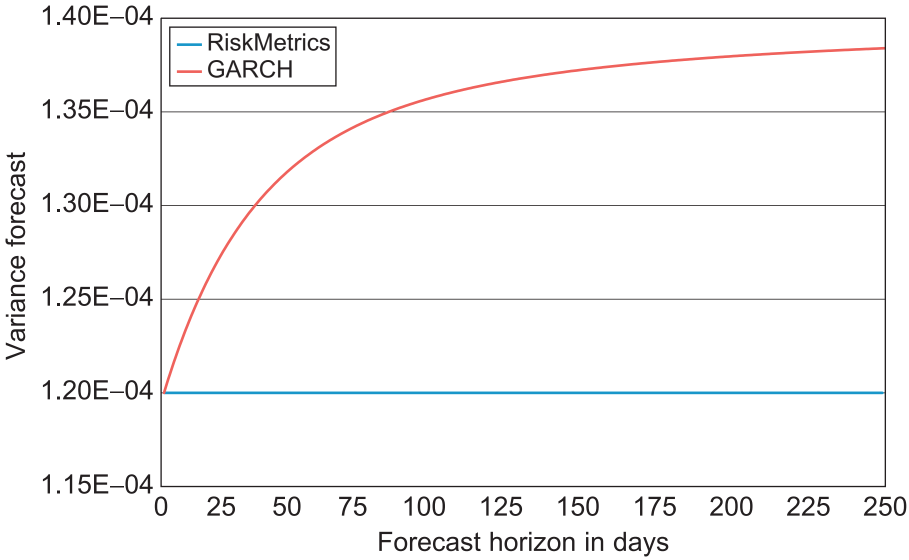
$$\sigma^2_{t+1:t+K} = \sum_{k=1}^{K} \sigma^2_{t+1} = K\sigma^2_{t+1}$$

But in the GARCH model, we get

$$\sigma^2_{t+1:t+K} = \sum_{k=1}^{K} \sigma^2_{t+1} = K\sigma^2 + \sum_{k=1}^{K} (\alpha + \beta)^{k-1}\left(\sigma^2_{t+1} - \sigma^2\right) \neq K\sigma^2_{t+1}$$

If the RiskMetrics and GARCH model have identical σ^2_{t+1}, and if $\sigma^2_{t+1} < \sigma^2$, then the GARCH variance forecast will be higher than the RiskMetrics forecast. Thus, assuming the RiskMetrics model if the data truly look more like GARCH will give risk managers a false sense of the calmness of the market in the future, when the market is calm today and $\sigma^2_{t+1} < \sigma^2$. Figure 4.2 illustrates this crucial point. We plot $\sigma^2_{t+1:t+K}/K$ for $K = 1, 2, \ldots, 250$ for both the RiskMetrics and the GARCH model starting from a low σ^2_{t+1} and setting $\alpha = 0.05$ and $\beta = 0.90$. The long-run daily variance in the figure is $\sigma^2 = 0.000140$.

Figure 4.2 Variance forecast for 1–250 days cumulative returns.



Notes: Assuming a common and low initial variance the plot shows the variance forecast for 1 through 250 trading days when using the RiskMetrics and GARCH model, respectively. The GARCH forecast converges to a long-run forecast that is above the current variance.

The GARCH and RiskMetrics models share the inconvenience that the multiperiod distribution is unknown even if the one-day ahead distribution is assumed to be normal, as we do in this chapter. Thus, while it is easy to forecast longer-horizon variance in these models, it is not as easy to forecast the entire conditional distribution. We will return to this important issue in Chapter 8 since it is unfortunately often ignored in risk management.

4 Maximum Likelihood Estimation

In the previous section, we suggested a GARCH model that we argued should fit the data well, but it contains a number of unknown parameters that must be estimated. In doing so, we face the challenge that the conditional variance, σ_{t+1}^2, is an unobserved variable, which must itself be implicitly estimated along with the parameters of the model, for example, α, β, and ω.

4.1 Standard Maximum Likelihood Estimation

We will briefly discuss the method of maximum likelihood estimation, which can be used to find parameter values. Explicitly worked out examples are included in the answers to the empirical exercises contained on the web site.

Recall our assumption that

$$R_t = \sigma_t z_t, \quad \text{with } z_t \sim \text{i.i.d. } N(0, 1)$$

The assumption of i.i.d. normality implies that the probability, or the likelihood, l_t, of R_t is

$$l_t = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{R_t^2}{2\sigma_t^2}\right)$$

and thus the joint likelihood of our entire sample is

$$L = \prod_{t=1}^{T} l_t = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{R_t^2}{2\sigma_t^2}\right)$$

A natural way to choose parameters to fit the data is then to maximize the joint likelihood of our observed sample. Recall that maximizing the logarithm of a function is equivalent to maximizing the function itself since the logarithm is a monotone, increasing function. Maximizing the logarithm is convenient because it replaces products with sums. Thus, we choose parameters $(\alpha, \beta, \ldots)$, which solve

$$Max \ln L = Max \sum_{t=1}^{T} \ln(l_t) = Max \sum_{t=1}^{T} \left[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\left(\sigma_t^2\right) - \frac{1}{2}\frac{R_t^2}{\sigma_t^2}\right]$$

and we refer to the optimal parameters as maximum likelihood estimates (MLEs). Note that the first term in the likelihood function is just a constant and so independent of the parameters of the models. We can therefore equally well optimize

$$Max \sum_{t=1}^{T} \left[-\frac{1}{2}\ln\left(\sigma_t^2\right) - \frac{1}{2}\frac{R_t^2}{\sigma_t^2}\right] = Max\left[-\frac{1}{2}\left(\sum_{t=1}^{T}\ln\left(\sigma_t^2\right) + \frac{R_t^2}{\sigma_t^2}\right)\right]$$

The MLE approach has the desirable property that as the sample size T goes to infinity the parameter estimates converge to their true values and the variance of these estimates are the smallest possible. In reality we of course do not have an infinite past of data available. Even if we have a long time series, say, of daily returns on the S&P 500 index available, it is not clear that we should use all that data when estimating the parameters. Sometimes obvious structural breaks such as a new exchange rate arrangement or new rules regulating trading in a particular market can guide in the choice of sample length. But often the dates of these structural breaks are not obvious and the risk manager is left with having to weigh the benefits of a longer sample, which implies more precise estimates (assuming there are no breaks), and a shorter sample, which reduces the risk of estimating across a structural break. When estimating GARCH models, a fairly good general rule of thumb is to use at least the past $1,000$ daily observations and to update the estimate sample fairly frequently, say monthly.

4.2 Quasi-Maximum Likelihood Estimation

The skeptical reader will immediately protest that the MLEs rely on the conditional normal distribution assumption, which we argued in Chapter 1 is false. While this protest appears to be valid, a key result in econometrics says that even if the conditional distribution is not normal, MLE will yield estimates of the mean and variance parameters that converge to the true parameters, when the sample gets infinitely large as long as the mean and variance functions are properly specified. This convenient result establishes what is called quasi-maximum likelihood estimation (QMLE), referring to the use of normal MLE estimation even when the normal distribution assumption is false. Notice that QMLE buys us the freedom to worry about the conditional distribution later (in Chapter 6), but it does come at a price: The QMLE estimates will in general be less precise than those from MLE. Thus, we trade theoretical asymptotic parameter efficiency for practicality.

The operational aspects of parameter estimation will be discussed in the exercises following this chapter. Here we just point out one simple but useful trick, which is referred to as variance targeting. Recall that the simple GARCH model can be written as

$$\sigma_{t+1}^2 = \omega + \alpha R_t^2 + \beta\sigma_t^2 = (1-\alpha-\beta)\sigma^2 + \alpha R_t^2 + \beta\sigma_t^2$$

Thus, instead of estimating ω by MLE, we can simply set the long-run variance, σ^2, equal to the sample variance, which is easily estimated beforehand as

$$\sigma^2 = \frac{1}{T}\sum_{t=1}^{T} R_t^2$$

Variance targeting has the benefit of imposing the long-run variance estimate on the GARCH model directly. More important, it reduces the number of parameters to be estimated in the model by one. This typically makes estimation much easier.

4.3 An Example

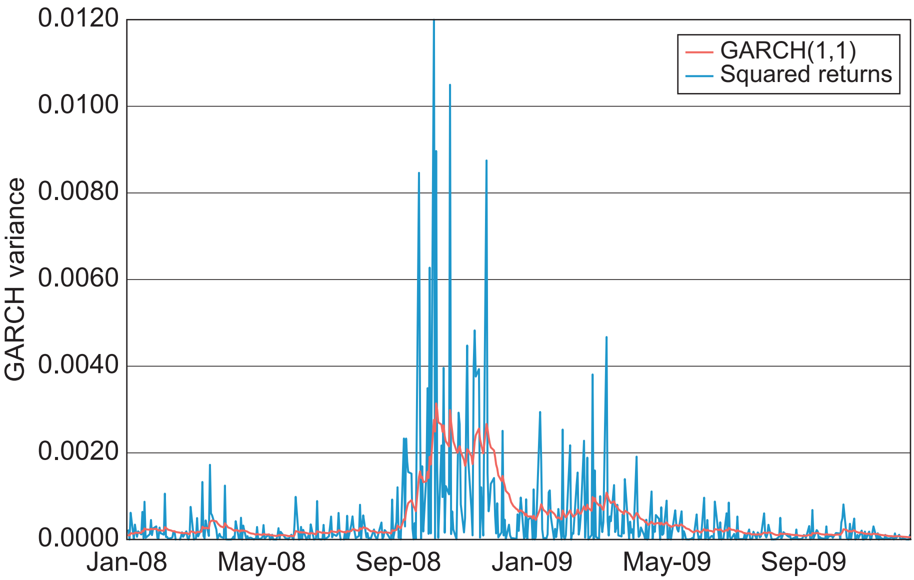
Figure 4.3 shows the S&P 500 squared returns from Figure 4.1, but now with an estimated GARCH variance superimposed. Using numerical optimization of the likelihood function (see the exercises at the end of the chapter), the optimal parameters imply the following variance dynamics:

$$\begin{aligned}\sigma_{t+1}^2 &= \omega + \alpha R_t^2 + \beta\sigma_t^2 \\ &= 0.0000011 + 0.100 \cdot R_t^2 + 0.899 \cdot \sigma_t^2\end{aligned}$$

The parameters have been estimated on the daily observations from January 1, 2001, through December 31, 2010. But we only plot the 2008–2009 period.

The persistence of variance in this model is $\alpha + \beta = 0.999$, which is only slightly lower than in RiskMetrics where it is 1. However, even if small, this difference

Figure 4.3 Squared S&P 500 returns with GARCH variance parameters estimated using QMLE.



Notes: The plot shows the daily squared return along with the daily GARCH variance forecast. The GARCH(1,1) model is used. The plots shows the 2008–2009 period.

will have consequences for the variance forecasts for horizons beyond one day. Furthermore, this very simple GARCH model may be misspecified driving the persistence close to one. We will consider more flexible models next.

5 Extensions to the GARCH Model

As we noted earlier, one of the distinct benefits of GARCH models is their flexibility. In this section, we explore this flexibility and present some of the models most useful for risk management.

5.1 The Leverage Effect

We argued in Chapter 1 that a negative return increases variance by more than a positive return of the same magnitude. This was referred to as the leverage effect, as a negative return on a stock implies a drop in the equity value, which implies that the company becomes more highly levered and thus more risky (assuming the level of debt stays constant). We can modify the GARCH models so that the weight given to the return depends on whether the return is positive or negative in the following simple

manner:

$$\sigma_{t+1}^2 = \omega + \alpha\left(R_t - \theta\sigma_t\right)^2 + \beta\sigma_t^2 = \omega + \alpha\sigma_t^2\left(z_t - \theta\right)^2 + \beta\sigma_t^2$$

which is sometimes referred to as the NGARCH (nonlinear GARCH) model.

Notice that it is strictly speaking a positive piece of news, $z_t > 0$, rather than raw return R_t, which has less of an impact on variance than a negative piece of news, if $\theta > 0$. The persistence of variance in this model is $\alpha(1+\theta^2)+\beta$, and the long-run variance is $\sigma^2 = \omega/(1-\alpha(1+\theta^2)-\beta)$.

Another way of capturing the leverage effect is to define an indicator variable, I_t, to take on the value 1 if day t's return is negative and zero otherwise.

$$I_t = \begin{cases} 1, & \text{if } R_t < 0 \\ 0, & \text{if } R_t \geq 0 \end{cases}$$

The variance dynamics can now be specified as

$$\sigma_{t+1}^2 = \omega + \alpha R_t^2 + \alpha\theta I_t R_t^2 + \beta\sigma_t^2$$

Thus, a θ larger than zero will capture the leverage effect. This is sometimes referred to as the GJR-GARCH model.

A different model that also captures the leverage is the exponential GARCH model or EGARCH,

$$\ln\sigma_{t+1}^2 = \omega + \alpha\left(\phi R_t + \gamma\left[|R_t| - E|R_t|\right]\right) + \beta\ln\sigma_t^2$$

which displays the usual leverage effect if $\alpha\phi < 0$. The EGARCH model has the advantage that the logarithmic specification ensures that variance is always positive, but it has the disadvantage that the future expected variance beyond one period cannot be calculated analytically.

5.2 More General News Impact Functions

Allowing for a leverage effect is just one way to extend the basic GARCH model. Many extensions are motivated by generalizing the way in which today's shock to return, z_t, impacts tomorrow's variance, σ_{t+1}^2. This relationship is referred to as the variance news impact function, *NIF*. In general we can write

$$\sigma_{t+1}^2 = \omega + \alpha\sigma_t^2 NIF(z_t) + \beta\sigma_t^2$$

In the simple GARCH model we have

$$NIF(z_t) = z_t^2$$

so that the news impact function is a symmetric parabola that takes the minimum value 0 when z_t is zero. In the NGARCH model with leverage we have

$$NIF(z_t) = (z_t - \theta)^2$$

so that the news impact function is still a parabola but now with the minimum value zero when $z_t = \theta$.

A very general news impact function can be defined by

$$NIF(z_t) = (|z_t - \theta_1| - \theta_2(z_t - \theta_1))^{2\theta_3}$$

Notice that the simple GARCH model is nested when $\theta_1 = \theta_2 = 0$, and $\theta_3 = 1$. The NGARCH model with leverage is nested when $\theta_2 = 0$, and $\theta_3 = 1$. The red lines in Figure 4.4 show the news impact function for various parameter values. The blue line in each panel denotes the simple symmetric GARCH model.

5.3 More General Dynamics

The simple GARCH model discussed earlier is often referred to as the GARCH(1,1) model because it relies on only one lag of returns squared and one lag of variance itself. For short-term variance forecasting, this model is often found to be sufficient, but in general we can allow for higher order dynamics by considering the GARCH(p,q) model, which simply allows for longer lags as follows:

$$\sigma_{t+1}^2 = \omega + \sum_{i=1}^{p} \alpha_i R_{t+1-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t+1-j}^2$$

These higher-order GARCH models have the disadvantage that the parameters are not easily interpretable. The component GARCH structure offers a great improvement in this regard. Let us go back to the GARCH(1,1) model to motivate earlier we can use $\sigma^2 = \omega/(1-\alpha-\beta)$ to rewrite the GARCH(1,1) model as

$$\sigma_{t+1}^2 = \sigma^2 + \alpha\left(R_t^2 - \sigma^2\right) + \beta\left(\sigma_t^2 - \sigma^2\right)$$

In the component GARCH model the long-run variance, σ^2, is allowed to be time varying and captured by the long-run variance factor v_{t+1}:

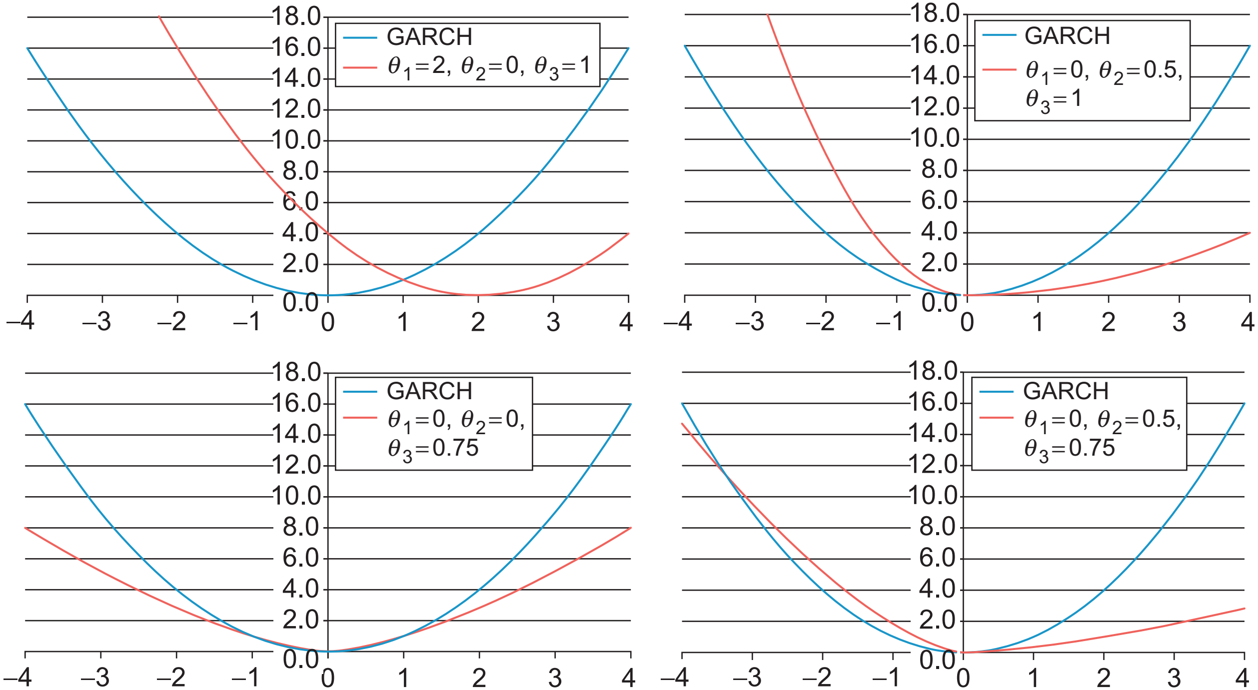
$$\sigma_{t+1}^2 = v_{t+1} + \alpha_\sigma\left(R_t^2 - v_t\right) + \beta_\sigma\left(\sigma_t^2 - v_t\right)$$
$$v_{t+1} = \sigma^2 + \alpha_v\left(R_t^2 - \sigma_t^2\right) + \beta_v\left(v_t - \sigma^2\right)$$

Note that the dynamic long-term variance, v_{t+1}, itself has a GARCH(1,1) structure. Thus we can think of the component GARCH model as being a GARCH(1,1) model around another GARCH(1,1) model.

The component model can potentially capture autocorrelation patterns in variance, which die out slower than what is possible in the simple shorter-memory GARCH(1,1) model. Appendix A shows that the component model can be rewritten as a GARCH(2,2) model:

$$\sigma_{t+1}^2 = \omega + \alpha_1 R_t^2 + \alpha_2 R_{t-1}^2 + \beta_1 \sigma_t^2 + \beta_2 \sigma_{t-1}^2$$

Figure 4.4 New impact functions.



Notes: The blue lines show the news impact function of the standard symmetric GARCH model. The red lines show the news impact function for various asymmetric GARCH models. The GARCH shock, z, is on the horizontal axis and the impact on variance from z (the news impact) is on the vertical axis.

where the parameters in the GARCH(2,2) are functions of the parameters in the component GARCH model.

But the component GARCH structure has the advantage that it is easier to interpret its parameters and therefore also easier to come up with good starting values for the parameters than in the GARCH(2,2) model. In the component model $\alpha_\sigma + \beta_\sigma$ capture the persistence of the short-run variance component and $\alpha_v + \beta_v$ capture the persistence in the long-run variance component. The GARCH(2,2) dynamic parameters α_1, α_2, β_1, β_2 have no such straightforward interpretation.

The component model dynamics can be extended with a leverage or an even more general news impact function as discussed earlier in the case of GARCH(1,1).

The HYGARCH model is a different model, which explicitly allows for long-memory, or hyperbolic decay, rather than exponential decay in variance. When modeling daily data, long memory in variance might often be relevant and it may be helpful in forecasting at longer horizons, say beyond a week. Consult Appendix B at the end of the chapter for information on the HYGARCH model, which is somewhat more complicated to implement.

5.4 Explanatory Variables

Because we are considering dynamic models of daily variance, we have to be careful with days where no trading takes place. It is widely recognized that days that follow a weekend or a holiday have higher variance than average days. As weekends and holidays are perfectly predictable, it makes sense to include them in the variance model. Other predetermined variables could be yesterday's trading volume or prescheduled news announcement dates such as company earnings and FOMC meetings dates. As these future events are known in advance, we can model

$$\sigma_{t+1}^2 = \omega + \beta\sigma_t^2 + \alpha\sigma_t^2 z_t^2 + \gamma IT_{t+1}$$

where IT_{t+1} takes on the value 1 if date $t+1$ is a Monday, for example.

We have not yet discussed option prices, but it is worth mentioning here that so-called implied volatilities from option prices often have quite high predictive value in forecasting next-day variance. Including the variance index (VIX) from the Chicago Board Options Exchange as an explanatory variable can improve the fit of a GARCH variance model of the underlying stock index significantly. Of course, not all underlying market variables have liquid options markets, so the implied volatility variable is not always available for variance forecasting. We will discuss the use of implied volatilities from options further in Chapter 10.

In general, we can write the GARCH variance forecasting model as follows:

$$\sigma_{t+1}^2 = \omega + h(X_t) + \alpha\sigma_t^2 z_t^2 + \beta\sigma_t^2$$

where X_t denotes variables known at the end of day t. As the variance is always a positive number, it is important to ensure that the GARCH model always generates a positive variance forecast. In the simple GARCH model, positive parameters ω, α

and β guarantee positivity. In the more general model considered here, positivity of $h(X_t)$—for example using the exponential function—along with positive ω, α, and β will ensure positivity of σ^2_{t+1}. We can write

$$\sigma^2_{t+1} = \omega + \exp\left(b'X_t\right) + \alpha\sigma^2_t z^2_t + \beta\sigma^2_t.$$

5.5 Generalizing the Low Frequency Variance Dynamics

These discussions of long memory and of explanatory variables motivates us to consider another extension to the simple GARCH models referred to as Spline-GARCH models. The daily variances path captured by the simple GARCH models is clearly itself very volatile. Volatility often spikes up for a few days and then quickly reverts back down to normal levels. Such quickly reverting spikes make volatility appear noisy and thus difficult to capture by explanatory variables. Explanatory variables may nevertheless be important for capturing longer-term trends in variance, which may thus need to be modeled separately so as to not be contaminated by the daily spikes.

In addition, some speculative prices may exhibit structural breaks in the level of volatility. A country may alter its foreign exchange regime, and a company may take over another company, for example, which could easily change its volatility structure.

In order to capture low-frequency changes in volatility we generalize the simple GARCH(1,1) model to the following multiplicative structure:

$$\begin{aligned}
\sigma^2_{t+1} &= \tau_{t+1} g_{t+1}, \text{ where} \\
g_{t+1} &= (1-\alpha-\beta) + \alpha g_t z^2_t + \beta g_t, \text{ and} \\
\tau_{t+1} &= \omega_0 \exp\left(\omega_1 t + \omega_2 \max\left(t - t_0, 0\right)^2 + \gamma X_t\right)
\end{aligned}$$

The Spline-GARCH model captures low frequency dynamics in variance via the τ_{t+1} process, and higher-frequency dynamics in variance via the g_{t+1} process. Notice that the low-frequency variance is kept positive via the exponential function. The low-frequency variance has a log linear time-trend captured by ω_1 and a quadratic time-trend starting at time t_0 and captured by ω_2. The low-frequency variance is also driven by the explanatory variables in the vector X_t.

Notice that the long-run variance in the Spline-GARCH model is captured by the low-frequency process

$$E\left[\sigma^2_{t+1}\right] = E\left[\tau_{t+1} g_{t+1}\right] = \tau_{t+1} E\left[g_{t+1}\right] = \tau_{t+1}$$

We can generalize the quadratic trend by allowing for many, say l, quadratic pieces, each starting at different time points and each with different slope parameters:

$$\tau_{t+1} = \omega_0 \exp\left(\omega_1 t + \sum_{i=1}^{l} \omega_{1+i} \max\left(t - t_{i-1}, 0\right)^2 + \gamma X_t\right)$$

Modeling equity index volatility at the country level, research has found that higher volatility is associated with lower market development (measured by market capitalization), larger GDP, higher inflation, higher GDP volatility, higher inflation volatility, and higher interest rate volatility.

5.6 Estimation of Extended Models

A particularly powerful feature of the GARCH family of models is that they can all be estimated using the same quasi MLE technique used for the simple GARCH(1,1) model. Regardless of the news impact functions, the dynamic structure, and the choice of explanatory variables, the model parameters can be estimated by maximizing the nontrivial part of the log likelihood

$$Max\left[-\frac{1}{2}\left(\sum_{t=1}^{T}\ln\left(\sigma_t^2\right)+\frac{R_t^2}{\sigma_t^2}\right)\right]$$

Notice as before that the variance path, σ_t^2, is a function of the parameters to be estimated.

6 Variance Model Evaluation

Before we start using the variance model for risk management purposes, it is appropriate to run the estimated model through some diagnostic checks.

6.1 Model Comparisons Using LR Tests

We have seen that the basic GARCH model can be extended by adding parameters and explanatory variables. The likelihood ratio test provides a simple way to judge if the added parameter(s) are significant in the statistical sense. Consider two different models with likelihood values L_0 and L_1, respectively. Assume that model 0 is a special case of model 1. In this case we can compare the two models via the likelihood ratio statistic

$$LR = 2\left(\ln\left(L_1\right) - \ln\left(L_0\right)\right)$$

The *LR* statistic will always be a positive number because model 1 contains model 0 as a special case and so model 1 will always fit the data better, even if only slightly so. The LR statistic will tell us if the improvement offered by model 1 over model 0 is statistically significant. It can be shown that the *LR* statistic will have a chi-squared distribution under the null hypothesis that the added parameters in model 1 are insignificant. If only one parameter is added then the degree of freedom in the chi-squared distribution will be 1. In this case the 1% critical value is approximately 6.63. A good rule of thumb is therefore that if the log-likelihood of model 1 is 3 to 4 points higher

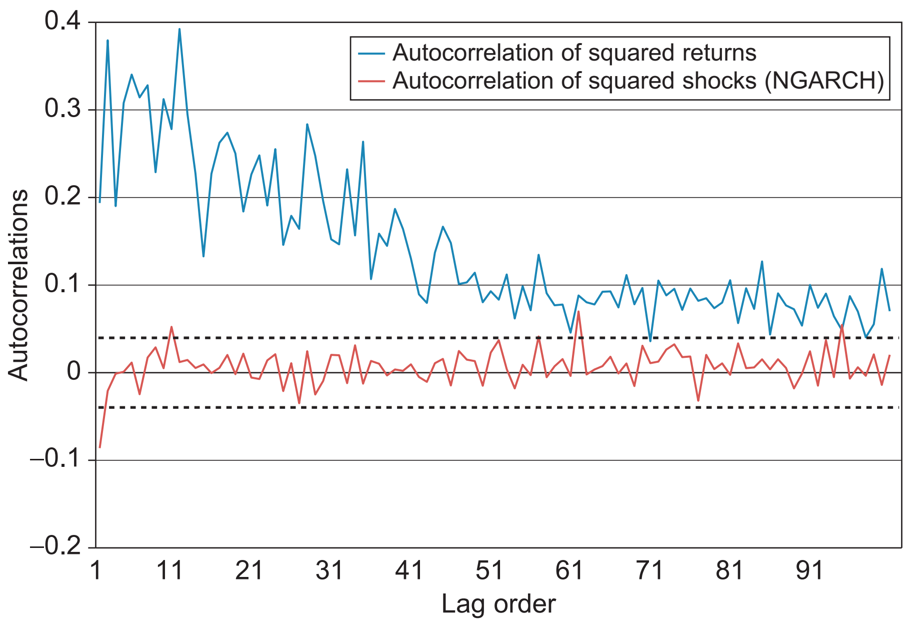
than that of model 0 then the added parameter in model 1 is significant. The degrees of freedom in the chi-squared test is equal to the number of parameters added in model 1.

6.2 Diagnostic Check on the Autocorrelations

In Chapter 1, we studied the behavior of the autocorrelation of returns and squared returns. We found that the raw return autocorrelations did not display any systematic patterns, whereas the squared return autocorrelations were positive for short lags and decreased as the lag order increased.

The objective of variance modeling is essentially to construct a variance measure, σ_t^2, which has the property that the standardized squared returns, R_t^2/σ_t^2, have no systematic autocorrelation patterns. Whether this has been achieved can be assessed via the red line in Figure 4.5, where we show the autocorrelation of R_t^2/σ_t^2 from the GARCH model with leverage for the S&P 500 returns along with their standard error bands. The standard errors are calculated simply as $1/\sqrt{T}$, where T is the number of observations in the sample. Usually the autocorrelation is shown along with plus/minus two standard error bands around zero, which simply mean horizontal lines at $-2/\sqrt{T}$ and $2/\sqrt{T}$. These so-called Bartlett standard error bands give the range in

Figure 4.5 Autocorrelation: Squared returns and squared returns over variance (NGARCH).



Notes: The blue line shows the autocorrelations of squared returns and the red line shows the autocorrelations of squared shocks defined as squared return divided by the GARCH variance. The NGARCH model is used here. The horizontal dashes denote 95% confidence bands.

which the autocorrelations would fall roughly 95% of the time if the true but unknown autocorrelations of R_t^2/σ_t^2 were all zero.

The blue line in Figure 4.5 redraws the autocorrelation of the squared returns from Chapter 1, now with the standard error bands superimposed. Comparing the two panels in Figure 4.5, we see that the GARCH model has been quite effective at removing the systematic patterns in the autocorrelation of the squared returns.

6.3 Volatility Forecast Evaluation Using Regression

Another traditional method of evaluating a variance model is based on simple regressions where squared returns in the forecast period, $t+1$, are regressed on the forecast from the variance model, as in

$$R_{t+1}^2 = b_0 + b_1\sigma_{t+1}^2 + e_{t+1}$$

A good variance forecast should be unbiased, that is, have an intercept $b_0 = 0$, and be efficient, that is, have a slope, $b_1 = 1$. In this regression, the squared return is used as a proxy for the true but unobserved variance in period $t+1$. One key question is, how good of a proxy is the squared return?

First of all, notice that it is true that $E_t[R_{t+1}^2] = \sigma_{t+1}^2$, so that the squared return is an unbiased proxy for true variance. But the variance of the proxy is

$$\begin{aligned} Var_t\left[R_{t+1}^2\right] &= E_t\left[\left(R_{t+1}^2 - \sigma_{t+1}^2\right)^2\right] = E_t\left[\left(\sigma_{t+1}^2(z_{t+1}^2 - 1)\right)^2\right] \\ &= \sigma_{t+1}^4 E_t\left[(z_{t+1}^2 - 1)^2\right] = \sigma_{t+1}^4(\kappa - 1) \end{aligned}$$

where κ is the kurtosis of the innovation, which is 3 under conditional normality but higher in reality. Thus, the squared return is an unbiased but potentially very noisy proxy for the conditional variance.

Due to the high degree of noise in the squared returns, the fit of the preceding regression as measured by the regression R^2 will be very low, typically around 5% to 10%, even if the variance model used to forecast is indeed the correct one. Thus, obtaining a low R^2 in such regressions should not lead us to reject the variance model. The conclusion is just as likely to be that the proxy for true but unobserved variance is simply very inaccurate.

In the next chapter we will look at ways to develop more accurate (realized volatility) proxies for daily variance using intraday return data.

6.4 The Volatility Forecast Loss Function

The ordinary least squares (OLS) estimation of a linear regression chooses the parameter values that minimize the mean squared error (MSE) in the regression. The regression-based approach to volatility forecast evaluation therefore implies a quadratic volatility forecast loss function. A correct volatility forecasting model

should have $b_0 = 0$ and $b_1 = 1$ as discussed earlier. A sensible loss function for comparing volatility models is therefore

$$\text{MSE} = \left(R_{t+1}^2 - \sigma_{t+1}^2\right)^2$$

Risk managers may however care differently about a negative versus a positive volatility forecast error of the same magnitude: Underestimating volatility may well be more costly for a conservative risk manager than overestimating volatility by the same amount.

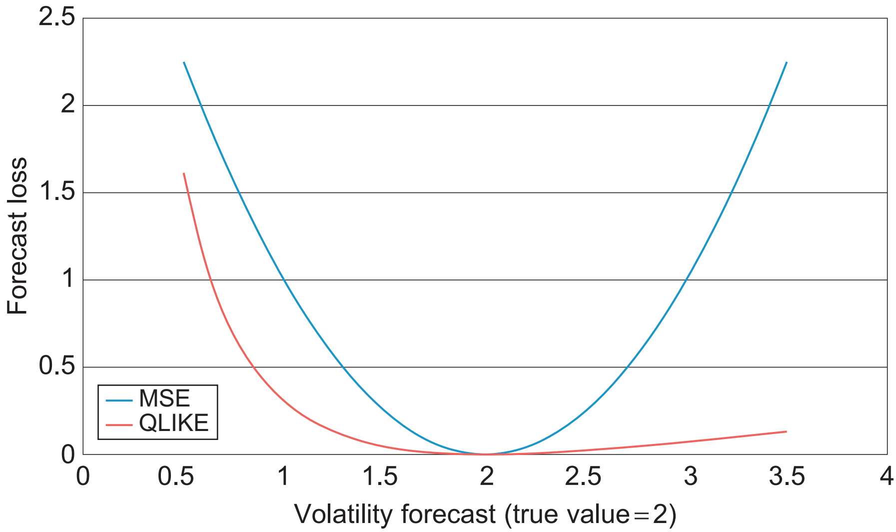
In order to evaluate volatility forecasts allowing for asymmetric loss, the following function can be used instead of MSE:

$$\text{QLIKE} = \frac{R_{t+1}^2}{\sigma_{t+1}^2} - \ln\left(\frac{R_{t+1}^2}{\sigma_{t+1}^2}\right) - 1$$

Notice that the QLIKE loss function depends on the relative volatility forecast error, R_{t+1}^2/σ_{t+1}^2, rather than on the absolute error, $|R_{t+1}^2 - \sigma_{t+1}^2|$; which is the key ingredient in MSE. The QLIKE loss function will always penalize more heavily volatility forecasts that underestimate volatility.

In Figure 4.6 we plot the MSE and QLIKE loss functions when the true volatility is 2 and the volatility forecast ranges from 0 to 4. Note the strong asymmetry of QLIKE for negative versus positive volatility forecast errors.

Figure 4.6 Volatility loss functions.



Notes: The blue line shows the mean squared error, MSE, loss function that is used in OLS and the red line shows the QLIKE loss function that is based on the likelihood function.

It can be shown that the MSE and the QLIKE loss functions are both robust with respect to the choice of volatility proxy. A volatility forecast loss function is said to be robust if the ranking of any two volatility forecasts is the same when using an observed volatility proxy as when (hypothetically) using the unobserved true volatility. Robustness is clearly a desirable quality for a volatility forecast loss function. Essentially only the MSE and QLIKE functions possess this quality.

Note again that we have only considered the noisy squared return, R_{t+1}^2, as a volatility proxy. In the next chapter we will consider more precise proxies using intraday information.

7 Summary

This chapter presented a range of variance models that are useful for risk management. Simple equally weighted and exponentially weighted models that require minimal effort in estimation were first introduced. Their shortcomings led us to consider more sophisticated but still simple models from the GARCH family. We highlighted the flexibility of GARCH as a virtue and considered various extensions to account for leverage effects, day-of-week effects, announcement effects, and so on. The powerful and flexible quasi-maximum likelihood estimation technique was presented and will be used again in coming chapters. Various model validation techniques were introduced subsequently. Most of the techniques suggested in this chapter are put to use in the empirical exercises that follow.

This chapter focused on models where daily returns are used to forecast daily volatility. In some situations the risk manager may have intraday returns available when forecasting daily volatility. In this case more accurate models are available. These models will be analyzed in the next chapter.

Appendix A: Component GARCH and GARCH(2,2)

This appendix shows that the component GARCH model can be viewed as a GARCH(2,2) model. First define the lag operator, L, to be a function that transforms the current value of a time series to the lagged value, that is,

$$Lx_t = x_{t-1}$$

Recall that the component GARCH model is given by

$$\sigma_{t+1}^2 = v_{t+1} + \alpha_\sigma \left(R_t^2 - v_t\right) + \beta_\sigma \left(\sigma_t^2 - v_t\right)$$

$$v_{t+1} = \sigma^2 + \alpha_v \left(R_t^2 - \sigma_t^2\right) + \beta_v \left(v_t - \sigma^2\right)$$

Using the lag operator, the first equation can be written as

$$\begin{aligned}\sigma_{t+1}^2 &= v_{t+1} + \alpha_\sigma \left(R_t^2 - Lv_{t+1}\right) + \beta_\sigma \left(\sigma_t^2 - Lv_{t+1}\right)\\ &= \alpha_\sigma R_t^2 + \beta_\sigma \sigma_t^2 + (1-(\alpha_\sigma+\beta_\sigma)L)v_{t+1}\end{aligned}$$

We can also rewrite the long-run variance equation using the lag operator

$$\begin{aligned}v_{t+1} &= \sigma^2 + \alpha_v \left(R_t^2 - \sigma_t^2\right) + \beta_v \left(v_t - \sigma^2\right)\\ &= \sigma^2 + \alpha_v R_t^2 - \alpha_v \sigma_t^2 + \beta_v L v_{t+1} - \beta_v \sigma^2\end{aligned}$$

From this, the long run variance factor v_{t+1} can be written as

$$v_{t+1} = \frac{1}{1-\beta_v L}\left(\left(1-\beta_v\right)\sigma^2 + \alpha_v R_t^2 - \alpha_v \sigma_t^2\right)$$

Plugging this back into the equation for σ_{t+1}^2 we get

$$\sigma_{t+1}^2 = \alpha_\sigma R_t^2 + \beta_\sigma \sigma_t^2 + \frac{(1-(\alpha_\sigma+\beta_\sigma)L)}{1-\beta_v L}\left((1-\beta_v)\sigma^2 + \alpha_v R_t^2 - \alpha_v \sigma_t^2\right)$$

Multiplying both sides by $1-\beta_v L$ and simplifying we get

$$\begin{aligned}\sigma_{t+1}^2 &= \beta_v L \sigma_{t+1}^2 + (1-\beta_v L)(\alpha_\sigma R_t^2 + \beta_\sigma \sigma_t^2)\\ &\quad + (1-(\alpha_\sigma+\beta_\sigma)L)\left((1-\beta_v)\sigma^2 + \alpha_v R_t^2 - \alpha_v \sigma_t^2\right)\\ &= \beta_v \sigma_t^2 + \alpha_\sigma R_t^2 + \beta_\sigma \sigma_t^2 - \beta_v \alpha_\sigma R_{t-1}^2 - \beta_v \beta_\sigma \sigma_{t-1}^2\\ &\quad + (1-(\alpha_\sigma+\beta_\sigma))(1-\beta_v)\sigma^2 + \alpha_v R_t^2\\ &\quad - (\alpha_\sigma+\beta_\sigma)\alpha_v R_{t-1}^2 - \alpha_v \sigma_t^2 + (\alpha_\sigma+\beta_\sigma)\alpha_v \sigma_{t-1}^2\\ &= (1-(\alpha_\sigma+\beta_\sigma))(1-\beta_v)\sigma^2 + (-\alpha_v+\beta_v+\beta_\sigma)\sigma_t^2\\ &\quad + ((\alpha_\sigma+\beta_\sigma)\alpha_v - \beta_v\beta_\sigma)\sigma_{t-1}^2\\ &\quad + (\alpha_\sigma+\alpha_v)R_t^2 + (-(\alpha_\sigma+\beta_\sigma)\alpha_v - \beta_v\alpha_\sigma)R_{t-1}^2\end{aligned}$$

which is a GARCH(2,2) model of the form

$$\sigma_{t+1}^2 = \omega + \alpha_1 R_t^2 + \alpha_2 R_{t-1}^2 + \beta_1 \sigma_t^2 + \beta_2 \sigma_{t-1}^2$$

where

$$\begin{aligned}\omega &= (1-(\alpha_\sigma+\beta_\sigma))(1-\beta_v)\sigma^2\\ \alpha_1 &= \alpha_\sigma + \alpha_v\\ \alpha_2 &= -(\alpha_\sigma+\beta_\sigma)\alpha_v - \beta_v\alpha_\sigma\\ \beta_1 &= -\alpha_v + \beta_v + \beta_\sigma\\ \beta_2 &= (\alpha_\sigma+\beta_\sigma)\alpha_v - \beta_v\beta_\sigma\end{aligned}$$

While the component GARCH and the GARCH(2,2) models are theoretically equivalent, the component structure is typically easier to implement because it is easier to find good starting values for the parameters. Typically, $\alpha_\sigma < \alpha_v \approx 0.1$ and $\beta_\sigma < \beta_v \approx 0.85$. Furthermore, the interpretation of v_{t+1} as a long-run variance when $(\alpha_v + \beta_v) > (\alpha_\sigma + \beta_\sigma)$ is convenient.

Appendix B: The HYGARCH Long-Memory Model

This appendix introduces the long-memory GARCH model of Davidson (2004), which he refers to as HYGARCH (where HY denotes hyperbolic). Consider first the standard GARCH(1,1) model where

$$\sigma_t^2 = \omega + \alpha R_{t-1}^2 + \beta \sigma_{t-1}^2$$

Using the lag operator, L, from Appendix A the GARCH(1,1) model can be written as an ARMA-in-squares model

$$(1 - (\alpha + \beta) L) R_t^2 = \omega + (1 - \beta L) \left(R_t^2 - \sigma_t^2 \right)$$

Solving for σ_t^2 we can now rewrite the GARCH(1,1) model as a moving-average-in-squares model of the form

$$\sigma_t^2 = \frac{\omega}{(1-\beta)} + \left(1 - \frac{1 - (\alpha + \beta) L}{1 - \beta L} \right) R_t^2$$

The long-memory model in Davidson (2004) is defined as a generalization of this expression to

$$\sigma_t^2 = \frac{\omega_d}{1 - \beta_d} + \left(1 - \frac{1 - \rho_d L}{1 - \beta_d L} \left(1 + \delta_d \left((1-L)^d - 1 \right) \right) \right) R_t^2$$

where the long-memory parameter, d, defines fractional integration via

$$(1-L)^d = 1 - \sum_{j=1}^{\infty} a_j L^j, \quad \text{with } a_j = \frac{d \Gamma(j-d)}{\Gamma(1-d) \Gamma(j+1)}$$

where $\Gamma(\bullet)$ denotes the gamma function. The long-memory parameter d enables the model to have squared return autocorrelations that go to zero at a slower (namely hyperbolic) rate than the exponential rate in the GARCH(1,1) case.

We can show that when $\rho_d = \alpha + \beta$, $\delta_d = 0$, $\omega_d = \omega$, and $\beta_d = \beta$ we get the regular GARCH(1,1) model as a special case of the HYGARCH. When $\rho_d = 1$, $\delta_d = 0$, $\omega_d = 0$, and $\beta_d = \lambda$ we get the RiskMetrics model as a special case of HYGARCH.

The HYGARCH model can be estimated using MLE as well. Please refer to Davidson (2004) for more details on this model.

Further Resources

The literature on variance modeling has exploded during the past 30 years, and we only present a few examples of papers here. Andersen et al. (2007) contains a discussion of the use of volatility models in risk management. Evidence on the performance of GARCH models through the 2008–2009 financial crisis can be found in Brownlees et al. (2009).

Andersen et al. (2006) and Poon and Granger (2005) provide overviews of the various classes of volatility models. The exponential smoother variance model is studied in JP Morgan (1996). The exponential smoother has been used to forecast a wide range of variables and further discussion of it can be found in Granger and Newbold (1986). The basic GARCH model is introduced in Engle (1982) and Bollerslev (1986), and it is discussed further in Bollerslev et al. (1992), and Engle and Patton (2001). Engle (2002) introduces the GARCH model with exogenous variable. Bollerslev (2008) summarizes the many variations on the basic GARCH model.

Long memory including fractional integrated generalized autoregressive conditional heteroskedasticity (FIGARCH) models were introduced in Baillie et al. (1996), and Bollerslev and Mikkelsen (1999). The hyperbolic HYGARCH model in Appendix B is developed in Davidson (2004).

Component volatility models were introduced in Engle and Lee (1999). They have been applied to option valuation in Christoffersen et al. (2010a, 2008). The Spline-GARCH model with a deterministic volatility component was introduced in Engle and Rangel (2008).

The leverage effect and other GARCH extensions are described in Ding et al. (1993), Glosten et al. (1993), Hentschel (1995), and Nelson (1990). Most GARCH models use squared return as the innovation to volatility. Forsberg and Ghysels (2007) analyze the benefits of using absolute returns instead.

Quasi maximum likelihood estimation of GARCH models is developed in Bollerslev and Wooldridge (1992). Francq and Zakoian (2009) survey more recent results. For practical issues involved in GARCH estimation see Zivot (2009). Hansen and Lunde (2005), Patton and Sheppard (2009), and Patton (2011) develop tools for volatility forecast evaluation and volatility forecast comparisons.

Taylor (1994) contains a very nice overview of a different class of variance models known as stochastic volatility models. This class of models was not included in this book due to the relative difficulty of estimating them.

References

Andersen, T., Christoffersen, P., Diebold, and F., 2006. Volatility and correlation forecasting. In: Elliott, G., Granger, C., Timmermann, A. (Eds.), The Handbook of Economic Forecasting. Elsevier, Amsterdam, The Netherlands, pp. 777–878.

Andersen, T., Bollerslev, T., Christoffersen, P., Diebold, F., 2007. Practical volatility and correlation modeling for financial market risk management. In: Carey, M., Stulz, R. (Eds.), The NBER Volume on Risks of Financial Institutions. University of Chicago Press, Chicago, IL, pp. 513–548.

Baillie, R., Bollerslev, T., Mikkelsen, H., 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. J. Econom. 74, 3–30.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econom. 31, 307–327.

Bollerslev, T., 2008, Glossary to ARCH (GARCH). Available from: SSRN, http://ssrn.com/abstract=1263250.

Bollerslev, T., Chou, R., Kroner, K., 1992. ARCH modeling in finance: A review of the theory and empirical evidence. J. Econom. 52, 5–59.

Bollerslev, T., Mikkelsen, H., 1999. Long-term equity anticipation securities and stock market volatility dynamics. J. Econom. 92, 75–99.

Bollerslev, T., Wooldridge, J., 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time varying covariances. Econom. Rev. 11, 143–172.

Brownlees, C., Engle, R., Kelly, B., 2009. A Practical Guide to Volatility Forecasting through Calm and Storm. Available from: SSRN, http://ssrn.com/abstract=1502915.

Christoffersen, P., Dorion, C., Jacobs, K., Wang, Y., 2010. Volatility components: Affine restrictions and non-normal innovations. J. Bus. Econ. Stat. 28, 483–502.

Christoffersen, P., Jacobs, K., Ornthanalai, C., Wang, Y., 2008. Option valuation with long-run and short-run volatility components. J. Financial Econ. 90, 272–297.

Davidson, J., 2004. Moment and memory properties of linear conditional heteroskedasticity models, and a new model. J. Bus. Econ. Stat. 22, 16–29.

Ding, Z., Granger, C.W.J., Engle, R.F., 1993. A long memory property of stock market returns and a new model. J. Empir. Finance 1(1), 83–106.

Engle, R., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. Econometrica 50, 987–1008.

Engle, R., 2002. New frontiers in ARCH models. J. Appl. Econom. 17, 425–446.

Engle, R., Lee, G., 1999. A permanent and transitory component model of stock return volatility. In: Engle, R., White, H. (Eds.), Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger, Oxford University Press, New York, NY, pp. 475–497.

Engle, R., Patton, A., 2001. What good is a volatility model? Quant. Finance 1, 237–245.

Engle, R., Rangel, J., 2008. The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. Rev. Financ. Stud. 21, 1187–1222.

Francq, C., Zakoian, J.-M., 2009. A tour in the asymptotic theory of GARCH estimation. In: Andersen, T.G., Davis, R.A., Kreib, J.-P., Mikosch, Th. (Eds.), Handbook of Financial Time Series, Heidelberg, Germany, pp. 85–112.

Forsberg, L., Ghysels, E., 2007. Why do absolute returns predict volatility so well? J. Financ. Econom. 5, 31–67.

Glosten, L., Jagannathan, R., Runkle, D., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. J. Finance 48, 1779–1801.

Granger, C.W.J., Newbold, P., 1986. Forecasting Economic Time Series, second ed. Academic Press, New York.

Hansen, P., Lunde, A., 2005. A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? J. Appl. Econom. 20, 873–889.

Hentschel, L., 1995. All in the family: Nesting symmetric and asymmetric GARCH models. J. Financ. Econ. 39, 71–104.
Morgan, J.P., 1996. RiskMetrics-Technical Document, fourth ed. J.P. Morgan/Reuters, New York, NY.
Nelson, D., 1990. Conditional heteroskedasticity in asset pricing: A new approach. Econometrica 59, 347–370.
Patton, A., 2011. Volatility forecast comparison using imperfect volatility proxies. J. Econom. 160, 246–256.
Patton, A., Sheppard, K., 2009. Evaluating volatility and correlation forecasts. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P., Mikosch, T. (Eds.), Handbook of Financial Time Series. Springer Verlag, Berlin.
Poon, S., Granger, C., 2005. Practical issues in forecasting volatility. Financ. Analyst. J. 61, 45–56.
Taylor, S., 1994. Modelling stochastic volatility: A review and comparative study. Math. Finance 4, 183–204.
Zivot, E., 2009. Practical issues in the analysis of univariate GARCH models. In: Andersen, T.G., Davis, R.A., Kreib, J.-P., Mikosch, Th. (Eds.), Handbook of Financial Time Series, Springer Verlag, Berlin, pp. 113–156.

Empirical Exercises

Open the Chapter4Data.xlsx file from the companion site.

A number of the exercises in this and the coming chapters rely on the maximum likelihood estimation (MLE) technique. The general approach to answering these questions is to use the parameter starting values to calculate the log likelihood value of each observation and then compute the sum of these individual log likelihoods. When using Excel, the Solver tool is then activated to maximize the sum of the log likelihoods by changing the cells corresponding to the parameter values. Solver is enabled through the Windows Office menu by selecting Excel Options and Add-Ins. When using Solver, choose the options Use Automatic Scaling and Assume Non-Negative. Set Precision, Tolerance, and Convergence to 0.0000001.

1. Estimate the simple GARCH(1,1) model on the S&P 500 daily log returns using the maximum likelihood estimation (MLE) technique. First estimate

$$\sigma_{t+1}^2 = \omega + \alpha R_t^2 + \beta\sigma_t^2, \quad \text{with } R_t = \sigma_t z_t, \quad \text{and } z_t \sim N(0,1)$$

Let the variance of the first observation be equal to the unconditional variance, $Var(R_t)$. Set the starting values of the parameters to $\alpha = 0.1$, $\beta = 0.85$, and $\omega = Var(R_t)(1-\alpha-\beta) \approx 0.01^2 \cdot 0.05 = 0.000005$. Reestimate the equation using variance targeting; that is, set $\omega = Var(R_t)(1-\alpha-\beta)$, and use Solver to find α and β only. Check how the estimated parameters and persistence differ from the variance model in Chapter 1.

2. Include a leverage effect in the variance equation. Estimate

$$\sigma_{t+1}^2 = \omega + \alpha\left(R_t - \theta\sigma_t\right)^2 + \beta\sigma_t^2, \quad \text{with } R_t = \sigma_t z_t, \quad \text{and } z_t \sim N(0,1)$$

Set starting values to $\alpha = 0.07$, $\beta = 0.85$, $\omega = 0.000005$, and $\theta = 0.5$. What is the sign of the leverage parameter? Explain how the leverage effect is captured in this model. Plot the

autocorrelations for lag 1 through 100 for R_t^2 as well as R_t^2/σ_t^2, and compare the two. Compare your results with Figure 4.5. Use an *LR* test to assess the significance of the leverage effect parameter θ.

3. Include the option implied volatility *VIX* series from the Chicago Board Options Exchange (CBOE) as an explanatory variable in the GARCH equation. Use MLE to estimate

$$\sigma_{t+1}^2 = \omega + \alpha\left(R_t - \theta\sigma_t\right)^2 + \beta\sigma_t^2 + \gamma VIX_t^2/252, \quad \text{with } R_t = \sigma_t z_t, \quad \text{and } z_t \sim N(0,1)$$

Set starting values to $\alpha = 0.04$, $\beta = 0.5$, $\omega = 0.000005$, $\theta = 2$, and $\gamma = 0.07$.

4. Estimate the component GARCH model defined by

$$\sigma_{t+1}^2 = v_{t+1} + \alpha_\sigma\left(R_t^2 - v_t\right) + \beta_\sigma\left(\sigma_t^2 - v_t\right)$$
$$v_{t+1} = \sigma^2 + \alpha_v\left(R_t^2 - \sigma_t^2\right) + \beta_v\left(v_t - \sigma^2\right)$$

The answers to these exercises can be found in the Chapter4Results.xlsx file on the companion site.

For more information see the companion site at
http://www.elsevierdirect.com/companions/9780123744487

5 Volatility Modeling Using Intraday Data

1 Chapter Overview

The goal of this chapter is to harness the information in intraday prices for computing daily volatility. Consider first estimating the mean of returns using a long sample of daily observations:

$$\hat{\mu} = \frac{1}{T}\sum_{t=0}^{T}\left(\ln\left(S_t\right) - \ln\left(S_{t-1}\right)\right) = \frac{1}{T}\left(\ln\left(S_T\right) - \ln\left(S_0\right)\right)$$

Note that when estimating the mean of returns only the first and the last observations matter: All the intermediate terms cancel out and their values are therefore completely inconsequential to the estimate of the mean. This result in turn implies that when estimating the mean, having a long time span of data is what matters: having daily versus weekly versus monthly data does not matter. The start and end points S_0 and S_T will be the same irrespective of the sampling frequency of returns. This is frustrating when we want to get a precise estimate of the return mean. The only solution is to wait for time to pass.

Consider now instead estimating variance on a sample of daily returns. We have

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=0}^{T}\left(\ln\left(S_t\right) - \ln\left(S_{t-1}\right) - \hat{\mu}\right)^2$$

Notice a crucial difference between the sample mean and sample variance estimators: The intermediate prices do not cancel out in the variance estimator. All the return observations now matter because they are squared before they are summed in the average.

Imagine now having price observations at the end of every hour instead of every day and imagine that the market for the asset at hand (for example an FX rate) is open 24 hours a day. Now we would have $24 \cdot T$ observations to estimate σ^2 and we would get a much more precise estimate than when using just the T daily returns.

The dramatic implication for risk management of this high-frequency sampling idea is that just as we can use 21 daily prices to estimate a monthly volatility we can

Elements of Financial Risk Management. DOI: 10.1016/B978-0-12-374448-7.00005-1

also use 24 hourly observations to estimate a daily volatility. If we have observations every minute then an even more precise estimate of daily volatility can be had and we can virtually treat daily volatility as an observed variable.

This chapter explores in detail the use of intraday prices for computing daily volatility and for forecasting future volatility. We first introduce the key concept of realized variance (RV) and look at four stylized facts of RV. We then look at ways to forecast RV and ways to estimate RV, and we briefly look at some of the challenges of working with large and messy intraday data sets. Toward the end of the chapter we look at range-based proxies of daily volatility and also at volatility forecast evaluation using RV and range-based volatility. Range-based volatilities are much easier to construct than RVs but in highly liquid markets RV will be more precise.

2 Realized Variance: Four Stylized Facts

Assume for simplicity that we are monitoring an asset that trades 24 hours per day and that is extremely liquid so that bid-ask spreads are virtually zero and new information is reflected in the price immediately. More realistic situations will be treated later. In an extremely liquid market with rapidly changing prices observed every second we can comfortably construct a time grid, for example, of 1-minute prices from which we can compute 1-minute log returns.

Let m be the number of observations per day on an asset. If we have 24 hour trading and 1-minute observations, then $m = 24 \cdot 60 = 1,440$. Let the jth observation on day $t+1$ be denoted $S_{t+j/m}$. Then the closing price on day $t+1$ is $S_{t+m/m} = S_{t+1}$, and the jth 1-minute return is

$$R_{t+j/m} = \ln(S_{t+j/m}) - \ln(S_{t+(j-1)/m})$$

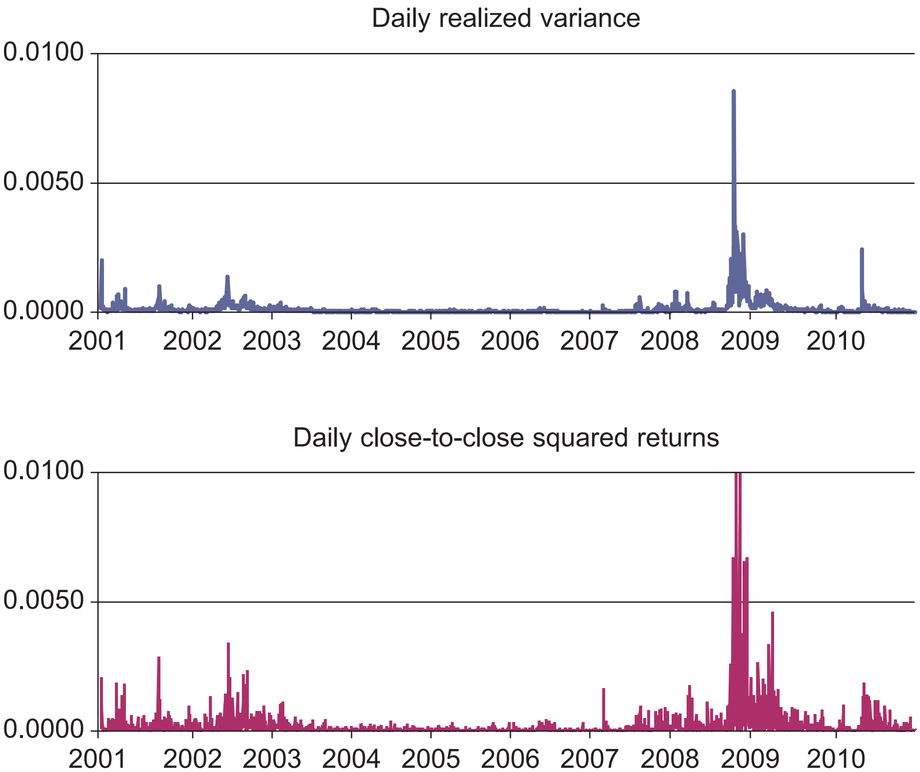
Having m observations available within a day, we can calculate an estimate of the daily variance from the intraday squared returns simply as

$$RV_{t+1}^m = \sum_{j=1}^{m} R_{t+j/m}^2$$

This is the definition of RV. Notice that unlike the previous chapters where we computed the sample variance from daily returns, we do not divide the sum of squared returns by m here. If we did we would get a 1-minute variance. Omitting the m gives us a total variance for the 24-hour period. Notice also that we do not subtract the mean of the 1-minute returns. The mean of 1-minute returns is so small that it will not materially impact the variance estimate.

The top panel of Figure 5.1 shows the time series of daily realized S&P 500 variance computed from intraday squared returns. The bottom panel shows the daily close-to-close squared returns S&P 500 as well. Notice how much more jagged and noisy

Figure 5.1 Realized variance (top) and squared returns (bottom) of the S&P 500.



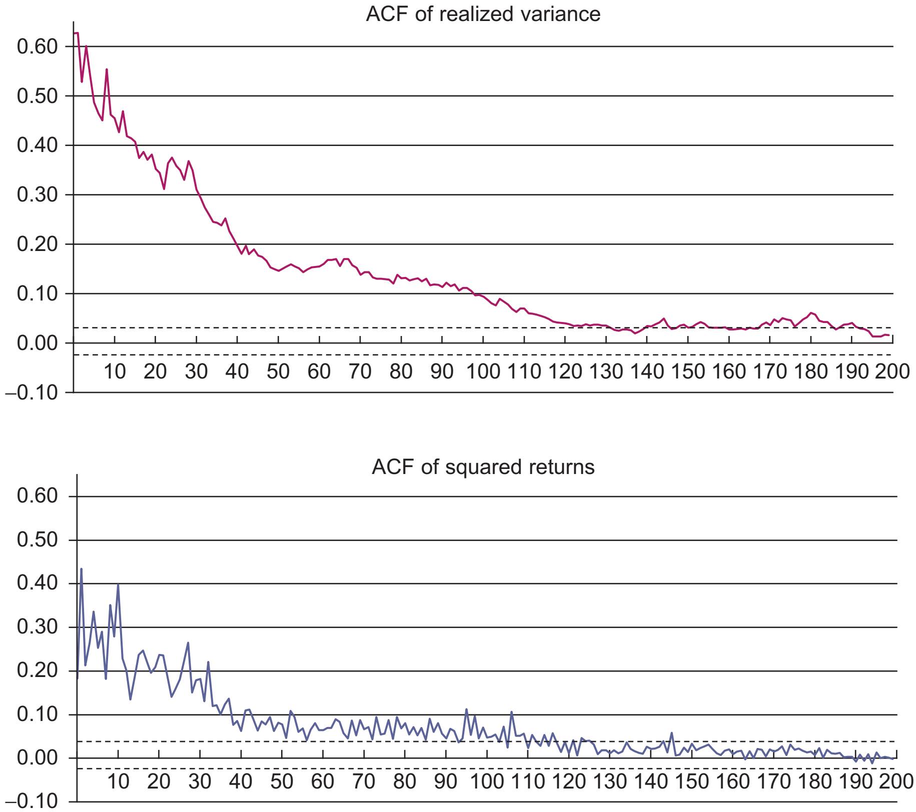
Notes: We use daily realized variance (top panel) and the daily close-to-close squared returns (bottom panel) as proxies for daily variance in the S&P 500 index.

the squared returns in the bottom panel are compared with the realized variances in the top panel. Figure 5.1 illustrates the first stylized fact of RV: RVs are much more precise indicators of daily variance than are daily squared returns.

The top panel of Figure 5.2 shows the autocorrelation function (ACF) of the S&P 500 RV series from Figure 5.1. The bottom panel shows the corresponding ACF computed from daily squared returns as in Chapter 4. Notice how much more striking the evidence of variance persistence is in the top panel. Figure 5.2 illustrates the second stylized fact of RV: RV is extremely persistent, which suggests that volatility may be forecastable at horizons beyond a few months as long as the information in intraday returns is used.

The top panel of Figure 5.3 shows a histogram of the RVs from Figure 5.1. The bottom panel of Figure 5.3 shows the histogram of the natural logarithm of RV. Figure 5.3 shows that the logarithm of RV is very close to normally distributed whereas the level of RV is strongly positively skewed with a long right tail.

Figure 5.2 Autocorrelation of realized variance (top) and autocorrelation of squared returns (bottom) with Bartlett confidence intervals (dashed).



Notes: We compute autocorrelations from the daily realized variance computed using the average RV method (top panel) and the daily close-to-close squared returns (bottom panel) from the S&P 500 index.

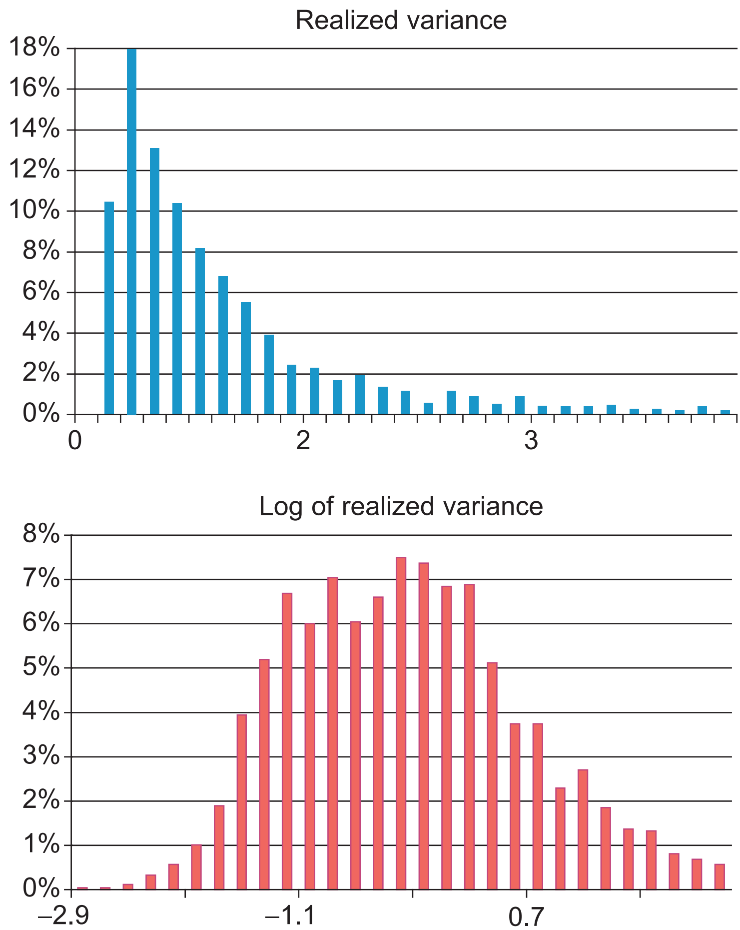
Given that RV is a sum of squared returns it is not surprising that RV is not close to normally distributed but it is interesting and useful that a simple logarithmic transformation results in a distribution that is somewhat close to normal. The approximate log normal property of RV is the third stylized fact. We can write

$$\ln\left(RV_{t+1}^{m}\right) \sim N\left(\mu_{RV}, \sigma_{RV}^{2}\right)$$

The fourth stylized fact of RV is that daily returns divided by the square root of RV is very close to following an i.i.d. (independently and identically distributed) standard normal distribution. We can write

$$R_{t+1}/\sqrt{RV_{t+1}^{m}} \overset{i.i.d.}{\sim} N(0,1)$$

Figure 5.3 Histogram of realized variance (top) and log realized variance (bottom).



Notes: We plot histograms of the daily realized variance computed using the average RV method (top panel) and the daily close-to-close squared returns (bottom panel) from the S&P 500 index.

Notice that because RV^m_{t+1} can only be computed at the end of day $t+1$ this result is not immediately useful for forecasting purposes.

The fourth stylized fact suggests that if a good forecast of RV^m_{t+1}, call it $RV^m_{t+1|t}$, can be made using information available at time t then a normal distribution assumption of $R_{t+1}/\sqrt{RV^m_{t+1|t}}$ will be a decent first modeling strategy. Approximately

$$R_{t+1}/\sqrt{RV^m_{t+1|t}} \overset{i.i.d.}{\sim} N(0,1)$$

where we have now standardized the return with the RV forecast, which by construction is known in advance. In this chapter we will rely on this assumption of normality

for the returns standardized by the RV forecast. In the next chapter we will allow for more general distributions.

Constructing a good forecast for RV_{t+1}^m is the topic to which we now turn. When doing so we will need to keep in mind the four stylized facts of RV:

- RV is a more precise indicator of daily variance than is the daily squared return.
- RV has large positive autocorrelations for many lags.
- The log of RV is approximately normally distributed.
- The daily return divided by the square root of RV is close to i.i.d. standard normal.

3 Forecasting Realized Variance

Realized variances are very persistent and so the main task at hand is to consider forecasting models that allow for current RV to matter for future RV.

3.1 Simple ARMA Models of Realized Variance

In Chapter 3 we introduced the AR(1) model as a simple way to allow for persistence in a time series. If we treat the estimated RV_t^m as an observed time series, then we can assume the AR(1) forecasting model

$$RV_{t+1}^m = \phi_0 + \phi_1 RV_t^m + \varepsilon_{t+1}$$

where ε_{t+1} is assumed to be uncorrelated over time and have zero mean. The parameters ϕ_0 and ϕ_1 can easily be estimated using OLS. The one-day-ahead forecast of RV is then constructed as

$$RV_{t+1|t}^m = \phi_0 + \phi_1 RV_t^m$$

We are just showing the AR(1) model as an example. AR(2) or higher ordered AR models could, of course, be used as well.

Given that we observed in Figure 5.3 that the log of RV is close to normally distributed we may be better off modeling the RV in logs rather than levels. We can therefore assume

$$\ln\left(RV_{t+1}^m\right) = \phi_0 + \phi_1 \ln\left(RV_t^m\right) + \varepsilon_{t+1}, \quad \text{with } \varepsilon_{t+1} \overset{i.i.d.}{\sim} N(0, \sigma_\varepsilon^2)$$

The normal property of $\ln\left(RV_{t+1}^m\right)$ will make the OLS estimates of ϕ_0 and ϕ_1 better behaved than those in the AR(1) model for RV_{t+1}^m where the AR(1) errors, ε_{t+1}, are likely to have fat tails, which in turn yield noisy parameter estimates.

Because we have estimated it from intraday squared returns, the RV_{t+1}^m series is not truly an observed time series but it can be viewed as the true RV observed with a measurement error. If the true RV is AR(1) but we observed true RV plus an i.i.d.

measurement error then an ARMA(1,1) model is likely to provide a good fit to the observed RV. We can write

$$\ln\left(RV_{t+1}^{m}\right) = \phi_0 + \phi_1 \ln\left(RV_t^{m}\right) + \theta_1 \varepsilon_t + \varepsilon_{t+1}, \quad \text{with } \varepsilon_{t+1} \overset{i.i.d.}{\sim} N(0, \sigma_\varepsilon^2)$$

which due to the MA term must be estimated using maximum likelihood techniques.

Notice that these simple models are specified in logarithms, while for risk management purposes we are ultimately interested in forecasting the level of variance. As the exponential function is *not* linear, we have in the log RV model that

$$RV_{t+1|t}^{m} = E_t\left[RV_{t+1}\right] = E_t\left[\exp\left(\ln\left(RV_{t+1}^{m}\right)\right)\right] \neq \exp\left(E_t\left[\ln\left(RV_{t+1}^{m}\right)\right]\right)$$

and we therefore have to be careful when calculating the variance forecast.

From the assumption of normality of the error term we can use the result

$$\varepsilon_{t+1} \sim N(0, \sigma_\varepsilon^2) \Longrightarrow E[\exp(\varepsilon_{t+1})] = \exp(\sigma_\varepsilon^2/2)$$

In the AR(1) model the forecast for tomorrow is

$$\begin{aligned} RV_{t+1|t}^{m} &= E_t[\exp(\phi_0 + \phi_1 \ln RV_t^{m} + \varepsilon_{t+1})] \\ &= \exp(\phi_0 + \phi_1 \ln RV_t^{m}) E_t[\exp(\varepsilon_{t+1})] \\ &= \left(RV_t^{m}\right)^{\phi_1} \exp(\phi_0 + \sigma_\varepsilon^2/2) \end{aligned}$$

and for the ARMA(1,1) model we get

$$\begin{aligned} RV_{t+1|t}^{m} &= E_t[\exp(\phi_0 + \phi_1 \ln RV_t^{m} + \theta_1 \varepsilon_t + \varepsilon_{t+1})] \\ &= \exp(\phi_0 + \phi_1 \ln RV_t^{m} + \theta_1 \varepsilon_t) E_t[\exp(\varepsilon_{t+1})] \\ &= \left(RV_t^{m}\right)^{\phi_1} \exp(\phi_0 + \theta_1 \varepsilon_t + \sigma_\varepsilon^2/2) \end{aligned}$$

More sophisticated models such as long-memory (or fractionally integrated) ARMA models can be used to model realized variance. These models may yield better longer horizon variance forecasts than the short-memory ARMA models considered here. As a simple but powerful way to allow for more persistence in the variance forecasting model we next consider the so-called heterogeneous AR models.

3.2 Heterogeneous Autoregressions (HAR)

The question arises whether we can parsimoniously (that is with only few parameters) and easily (that is using OLS) model the apparent long-memory features of realized volatility. The mixed-frequency or heterogeneous autoregression model (HAR) we now consider provides an affirmative answer to this question. Define the h-day RV from the 1-day RV as follows:

$$RV_{t-h+1,t} = [RV_{t-h+1} + RV_{t-h+2} + \cdots + RV_t]/h$$

where dividing by h makes $RV_{t-h+1,t}$ interpretable as the average total variance starting with day $t-h+1$ and through day t.

Given that economic activity is organized in days, weeks, and months, it is natural to consider forecasting tomorrow's RV using daily, weekly, and monthly RV defined by the simple moving averages

$$\begin{aligned} RV_{D,t} &\equiv RV_t \\ RV_{W,t} &\equiv RV_{t-4,t} = [RV_{t-4} + RV_{t-3} + RV_{t-2} + RV_{t-1} + RV_t]/5 \\ RV_{M,t} &\equiv RV_{t-20,t} = [RV_{t-20} + RV_{t-19} + \cdots + RV_t]/21 \end{aligned}$$

where we have assumed five trading days in a week and 21 trading days in a month. The simplest way to forecast RV with these variables is via the regression

$$RV_{t+1} = \phi_0 + \phi_D RV_{D,t} + \phi_W RV_{W,t} + \phi_M RV_{M,t} + \varepsilon_{t+1}$$

which defines the HAR model. Notice that HAR can be estimated by OLS because all variables are observed and because the model is linear in the parameters.

The HAR will be able to capture long-memory-like dynamics because 21 lags of daily RV matter in this model. The model is parsimonious because the 21 lags of daily RV do not have 21 different autoregressive coefficients: The coefficients are restricted to be $\left(\phi_D + \phi_W/5 + \phi_M/21\right)$ on today's RV, $\left(\phi_W/5 + \phi_M/21\right)$ on the past four days of RV, and $\phi_M/21$ on the RVs for days $t-20$ through $t-5$.

Given the log normal property of RV we can also consider HAR models of the log transformation of RV:

$$\begin{aligned} \ln(RV_{t+1}) &= \phi_0 + \phi_D \ln(RV_{D,t}) + \phi_W \ln(RV_{W,t}) \\ &\quad + \phi_M \ln(RV_{M,t}) + \varepsilon_{t+1}, \quad \text{with } \varepsilon_{t+1} \stackrel{i.i.d.}{\sim} N(0, \sigma_\varepsilon^2) \end{aligned}$$

The advantage of this log specification is again that the parameters will be estimated more precisely when using OLS. Remember though that forecasting involves undoing the log transformation so that

$$\begin{aligned} RV^m_{t+1|t} &= \exp\left(\phi_0 + \phi_D \ln(RV_{D,t}) + \phi_W \ln(RV_{W,t}) + \phi_M \ln(RV_{M,t})\right) \exp(\sigma_\varepsilon^2/2) \\ &= (RV_t)^{\phi_D} \left(RV_{W,t}\right)^{\phi_W} \left(RV_{M,t}\right)^{\phi_M} \exp(\phi_0 + \sigma_\varepsilon^2/2) \end{aligned}$$

Note that the HAR idea generalizes to longer-horizon forecasting. If for example we want to forecast RV over the next K days then we can estimate the model

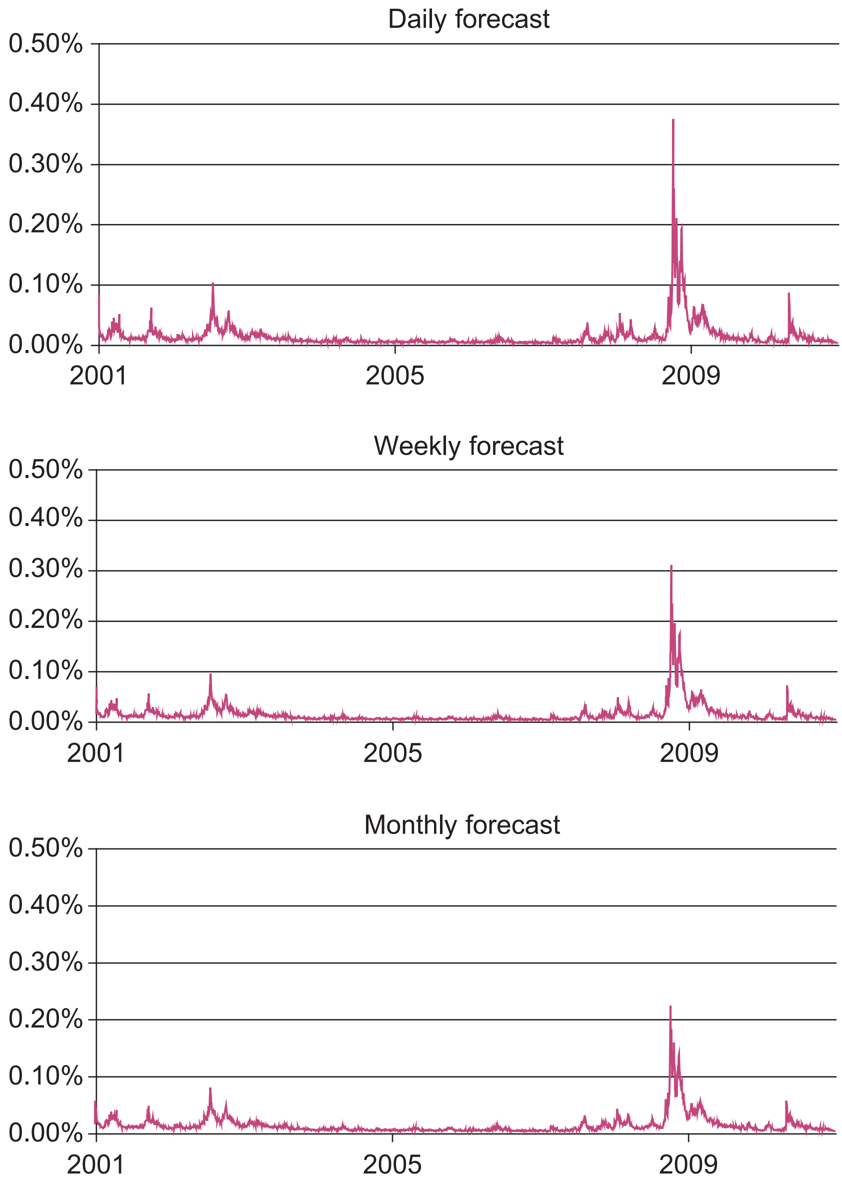
$$\begin{aligned} \ln(RV_{t+1,t+K}) &= \phi_{0,K} + \phi_{D,K} \ln(RV_{D,t}) + \phi_{W,K} \ln(RV_{W,t}) \\ &\quad + \phi_{M,K} \ln(RV_{M,t}) + \varepsilon_{t+1,t+K} \end{aligned}$$

where

$$RV_{t+1,t+K} = [RV_{t+1} + RV_{t+2} + \cdots + RV_{t+K}]/K$$

and where we still rely on daily, weekly, and monthly RVs on the right-hand side of the HAR model. Figure 5.4 shows the forecast of 1-day, 5-day, and 10-day volatility using three different log HAR models corresponding to each horizon of interest.

Figure 5.4 Forecast of daily (top), weekly (middle), and monthly (bottom) S&P 500 volatility using HAR model specified in logs.



Notes: We use the HAR model estimated in logs to forecast the level of variance over the next day, week, and month in the S&P 500 index.

In Chapter 4 we saw the importance of including a leverage effect in the GARCH model capturing that volatility rises more on a large positive return than on a large negative return. The HAR model can capture this by simply including the return on

the right-hand side. In the daily log HAR we can write

$$\ln(RV_{t+1}) = \phi_0 + \phi_D \ln(RV_{D,t}) + \phi_W \ln(RV_{W,t}) + \phi_M \ln(RV_{M,t}) + \phi_R R_t + \varepsilon_{t+1}$$

which can also easily be estimated using OLS. Notice that because the model is written in logs we do not have to worry about the variance forecast going negative;

$$RV_{t+1|t} = E_t\left[\exp\left(\ln\left(RV^m_{t+1}\right)\right)\right]$$

will always be a positive number.

The stylized facts of RV suggested that we can assume that

$$R_{t+1}/\sqrt{RV^m_{t+1|t}} \overset{i.i.d.}{\sim} N(0,1)$$

If we use this assumption then from Chapter 1 we can compute Value-at-Risk by

$$VaR^p_{t+1} = -\sqrt{RV^m_{t+1|t}}\Phi^{-1}_p$$

where $RV^m_{t+1|t}$ is provided by either the ARMA or HAR forecasting models earlier. Expected Shortfall is also easily computed via

$$ES^p_{t+1} = \sqrt{RV^m_{t+1|t}}\frac{\phi\left(\Phi^{-1}_p\right)}{p}$$

which follows from Chapter 2.

3.3 Combining GARCH and RV

So far in Chapters 4 and 5 we have considered two seemingly very different approaches to volatility modeling: In Chapter 4 GARCH models were estimated on daily returns, and in Chapter 5 time-series models of daily RV have been constructed from intraday returns. We can instead try to incorporate the rich information in RV into a GARCH modeling framework. Consider the basic GARCH model from Chapter 4:

$$R_{t+1} = \sigma_{t+1} z_{t+1}, \text{ where}$$
$$\sigma^2_{t+1} = \omega + \alpha R^2_t + \beta\sigma^2_t$$

Given the information on daily RV we could augment the GARCH model with RV as follows:

$$\sigma^2_{t+1} = \omega + \alpha R^2_t + \beta\sigma^2_t + \gamma RV^m_t$$

This so-called GARCH-X model where RV is the explanatory variable can be estimated using the univariate MLE approach taken in Chapter 4.

A shortcoming of the GARCH-X approach is that a model for RV is not specified. This means that we cannot use the model to forecast volatility beyond one day ahead.

The more general so-called Realized GARCH model is defined by

$$R_{t+1} = \sigma_{t+1} z_{t+1}, \text{ where}$$
$$\sigma^2_{t+1} = \omega + \alpha R^2_t + \beta \sigma^2_t + \gamma RV^m_t, \text{ and}$$
$$RV^m_t = \omega_{RV} + \beta_{RV} \sigma^2_t + \varepsilon_t$$

where ε_t is the innovation to RV. This model can be estimated by MLE when assuming that R_t and ε_t have a joint normal distribution. The Realized GARCH model can be augmented to include a leverage effect as well. In the Realized GARCH model the *VaR* and *ES* would simply be

$$VaR^p_{t+1} = -\sigma_{t+1} \Phi^{-1}_p$$

and

$$ES^p_{t+1} = \sigma_{t+1} \frac{\phi\left(\Phi^{-1}_p\right)}{p}$$

as in the regular GARCH model.

4 Realized Variance Construction

So far we have assumed that a grid of highly liquid 1-minute prices are available so that the corresponding 1-minute log returns are informative about the true volatility of the asset price. However, once various forms of illiquidity in the asset price are considered it becomes clear that we need to be much more clever about constructing the RVs from the intraday returns. This section is devoted to the construction of unbiased daily RVs from intraday returns under realistic assumptions about market liquidity.

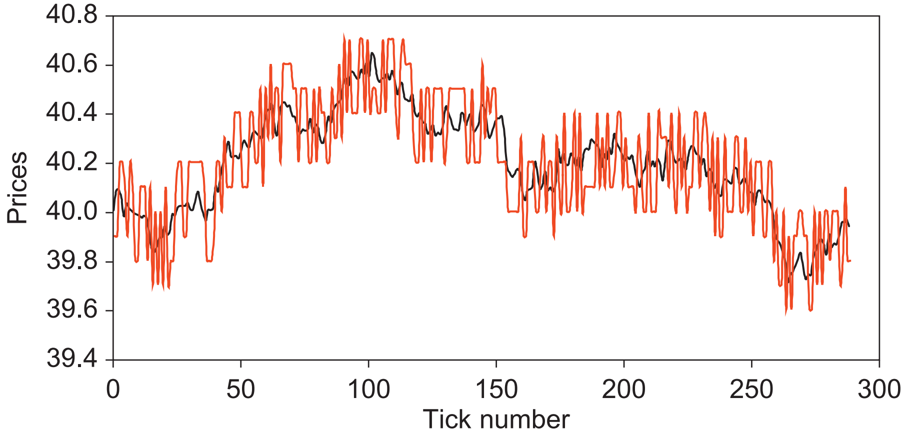
4.1 The All RV Estimator

Remember that in the ideal but unfortunately unrealistic case with ultra-high liquidity we have $m = 24 \cdot 60$ observations available within a day, and we can calculate an estimate of the daily variance from the intraday squared returns simply as

$$RV^m_{t+1} = \sum_{j=1}^{m} R^2_{t+j/m} = \sum_{j=1}^{m} \left(\ln(S_{t+j/m}) - \ln(S_{t+(j-1)/m})\right)^2$$

This estimator is sometimes known as the All RV estimator because it uses all the prices on the 1-minute grid.

Figure 5.5 Fundamental price and quoted price with bid-ask bounces.



Notes: We simulate a random walk for the fundamental log asset price (black) and add random noise from bid-ask bounces to get the observed price (red).

Figure 5.5 uses simulated data to illustrate one of the problems caused by illiquidity when estimating asset price volatility. We assume the fundamental (but unobserved) asset price, S^{Fund}, follows the simple random walk process with constant variance

$$\ln S^{Fund}_{t+j/m} = \ln S^{Fund}_{t+(j-1)/m} + e_{t+j/m}, \text{ with } e_{t+j/m} \overset{i.i.d.}{\sim} N(0, \sigma_e^2)$$

where $\sigma_e = 0.001$ in Figure 5.5. The observed price fluctuates randomly around the bid and ask quotes that are posted by the market maker. We observe

$$S^{Obs}_{t+j/m} = B_{t+j/m} I_{t+j/m} + A_{t+j/m}(1 - I_{t+j/m})$$

where $B_{t+j/m}$ is the bid price, which we take to be the fundamental price rounded down to the nearest \$1/10, and $A_{t+j/m}$ is the ask price, which is the fundamental price rounded up to the nearest \$1/10. $I_{t+j/m}$is an i.i.d. random variable, which takes the values 1 and 0 each with probability 1/2. $I_{t+j/m}$ is thus an indicator variable of whether the observed price is a bid or an ask price.

The challenge is that we observe $S^{Obs}_{t+j/m}$ but want to estimate σ^2, which is the variance of the unobserved $S^{Fund}_{t+j/m}$. Figure 5.5 shows that the observed intraday price can be very noisy compared with the smooth fundamental but unobserved price. The bid-ask spread adds a layer of noise on top of the fundamental price. If we compute RV^m_{t+1} from the high-frequency $S^{Obs}_{t+j/m}$ then we will get an estimate of σ^2 that is higher than the true value because of the inclusion of the bid-ask volatility in the estimate.

4.2 The Sparse RV Estimator

The perhaps simplest way to address the problem shown in Figure 5.5 is to construct a grid of intraday prices and returns that are sampled less frequently than the 1-minute assumed earlier. Instead of a 1-minute grid we could use an s-minute grid (where $s \geq 1$) so that our new RV estimator would be

$$RV^s_{t+1} = \sum_{j=1}^{m/s} R^2_{t+js/m}$$

which is sometimes denoted as the Sparse RV estimator as opposed to the previous All RV estimator.

Of course the important question is how to choose the parameter s? Should s be 5 minutes, 10 minutes, 30 minutes, or an even lower frequency? The larger the s the less likely we are to get a biased estimate of volatility, but the larger the s the fewer observations we are using and so the more noisy our estimate will be. We are faced with a typical variance-bias trade-off.

The choice of s clearly depends on the specific asset. For very liquid assets we should use an s close to 1 and for illiquid assets s should be much larger. If liquidity effects manifest themselves as a bias in the estimated RVs when using a high sampling frequency then that bias should disappear when the sampling frequency is lowered; that is, when s is increased.

The so-called volatility signature plots provide a convenient graphical tool for choosing s: First compute RV^s_{t+1} for values of s going from 1 to 120 minutes. Second, scatter plot the average RV across days on the vertical axis against s on the horizontal axis. Third, look for the smallest s such that the average RV does not change much for values of s larger than this number.

In markets with wide bid–ask spreads the average RV in the volatility signature plot will be downward sloping for small s but for larger s the average RV will stabilize at the true long run volatility level. We want to choose the smallest s for which the average RV is stable. This will avoid bias and minimize variance.

In markets where trading is thin, new information is only slowly incorporated into the price, and intraday returns will have positive autocorrelation resulting in an upward sloping volatility signature plot. In this case, the rule of thumb for computing RV is again to choose the smallest s for which the average RV has stabilized.

4.3 The Average RV Estimator

Choosing a lower (sparse) frequency for the grid of intraday prices can solve the bias problem arising from illiquidity but it will also increase the noise of the RV estimator. When we are using sparse sampling we are essentially throwing away information, which seems wasteful. It turns out that there is an amazingly simple way to lower the noise of the Sparse RV estimator without increasing the bias.

Let us say that we have used the volatility signature plot to chose $s = 15$ in the Sparse RV so that we are using a 15-minute grid for prices and squared returns to compute RV. Note that if we have the original 1-minute grid (of less liquid prices) then we can actually compute 15 different (but overlapping) Sparse RV estimators. The first Sparse RV will use a 15-minute grid starting with the 15-minute return at midnight, call it $RV_{t+1}^{s,1}$; the second will also use a 15-minute grid but this one will be starting one minute past midnight, call it $RV_{t+1}^{s,2}$, and so on until the 15th Sparse RV, which uses a 15-minute grid starting at 14 minutes past midnight, call it $RV_{t+1}^{s,15}$. We are thus using the fine 1-minute grid to compute 15 Sparse RVs at the 15-minute frequency.

We have now used all the information on the 1-minute grid but we have used it to compute 15 different RV estimates, each based on 15-minute returns, and none of which are materially affected by illiquidity bias. By simply averaging the 15 sparse RVs we get the so-called Average RV estimator

$$RV_{t+1}^{Avr} = \frac{1}{s}\sum_{i=1}^{s} RV_{t+1}^{s,i}$$

In simulation studies and in practice this Average RV estimator has been found to perform very well. The RVs plotted in Figure 5.1 were computed using the Average RV estimator.

4.4 RV Estimators with Autocovariance Adjustments

Instead of using sparse sampling to avoid RV bias we can try to model and then correct for the autocorrelations in intraday returns that are driving the volatility bias.

Assume that the fundamental log price is observed with an additive i.i.d. error term, u, caused by illiquidity so that

$$\ln(S_{t+j/m}^{Obs}) = \ln(S_{t+j/m}^{Fund}) + u_{t+j/m}, \text{ with } u_{t+j/m} \overset{i.i.d.}{\sim} N\left(0, \sigma_u^2\right)$$

In this case the observed log return will equal the true fundamental returns plus an MA(1) error:

$$\begin{aligned} R_{t+j/m}^{Obs} &= \ln(S_{t+j/m}^{Obs}) - \ln(S_{t+(j-1)/m}^{Obs}) \\ &= \ln(S_{t+j/m}^{Fund}) + u_{t+j/m} - \left(\ln(S_{t+(j-1)/m}^{Fund}) + u_{t+(j-1)/m}\right) \\ &= R_{t+j/m}^{Fund} + u_{t+j/m} - u_{t+(j-1)/m} \end{aligned}$$

Due to the MA(1) measurement error our simple squared return All RV estimate will be biased.

The All RV in this case is defined by

$$RV_{t+1}^{m} = \sum_{j=1}^{m}\left(R_{t+j/m}^{Obs}\right)^2 = \sum_{j=1}^{m}\left(R_{t+j/m}^{Fund} + u_{t+j/m} - u_{t+(j-1)/m}\right)^2$$

Because the measurement error u has positive variance the RV_{t+1}^{m} estimator will be biased upward in this case.

If we are fairly confident that the measurement error is of the MA(1) form then we know (see Chapter 3) that only the first-order autocorrelations are nonzero and we can therefore easily correct the RV estimator as follows:

$$RV_{t+1}^{AR(1)} = \sum_{j=1}^{m}\left(R_{t+j/m}^{Obs}\right)^2 + \sum_{j=2}^{m} R_{t+j/m}^{Obs} R_{t+(j-1)/m}^{Obs} + \sum_{j=1}^{m-1} R_{t+j/m}^{Obs} R_{t+(j+1)/m}^{Obs}$$

where we have added the cross products from the adjacent intraday returns. The negative autocorrelation arising from the bid–ask bounce in observed intraday returns will cause the last two terms in $RV_{t+1}^{AR(1)}$ to be negative and we will therefore get that

$$RV_{t+1}^{AR(1)} < RV_{t+1}^{m} = \sum_{j=1}^{m}\left(R_{t+j/m}^{Obs}\right)^2$$

as desired.

Positive autocorrelation caused by slowly changing prices would be at least partly captured by the first-order autocorrelation as well. It would be positive in this case and we would have

$$RV_{t+1}^{AR(1)} > RV_{t+1}^{m}$$

Much more general estimators have been developed to correct for more complex autocorrelation patterns in intraday returns. References to this work will be listed at the end of the chapter.

5 Data Issues

So far we have assumed the availability of a 1-minute grid of prices in a 24-hour market. But in reality several challenges arise. First, prices and quotes arrive randomly in time and not on a neat, evenly spaced grid. Second, markets are typically not open 24 hours per day. Third, intraday data sets are large and messy and often include price and quote errors that must be flagged and removed before estimating volatility. We deal with these three issues in turn as we continue.

5.1 Dealing with Irregularly Spaced Intraday Prices

The preceding discussion has assumed that a sample of regularly spaced 1-minute intraday prices are available. In practice, transaction prices or quotes arrive in random ticks over time and the evenly spaced price grid must be constructed from the raw ticks.

One of the following two methods are commonly used.

The first and simplest solution is to use the last tick prior to a grid point as the price observation for that grid point. This way the last observed tick price in an interval is effectively moved forward in time to the next grid point. Specifically, assume we have N observed tick prices during day $t+1$ but that these are observed at irregular times $t(0), t(1), \ldots, t(N)$. Consider now the jth point on the evenly spaced grid of m points for day $t+1$, which we have called $t+j/m$. Grid point $t+j/m$ will fall between two adjacent randomly spaced ticks, say the ith and the $(i+1)$th; that is, we have $t(i) < t+j/m < t(i+1)$ and in this case we choose the $t+j/m$ price to be

$$S_{t+j/m} = S_{t(i)}$$

The second and slightly less simple solution uses a linear interpolation between $S_{t(i)}$ and $S_{t(i+1)}$ so what we have

$$S_{t+j/m} = S_{t(i)} + \frac{(t+j/m) - t(i)}{t(i+1) - t(i)}[S_{t(i+1)} - S_{t(i)}], \quad \text{for } t(i) < t+j/m < t(i+1)$$

While the linear interpolation method makes some intuitive sense it has poor limiting properties: The smoothing implicit in the linear interpolation makes the estimated RV go to zero in the limit. Therefore, using the most recent tick on each grid point has become standard practice.

5.2 Choosing the Frequency of the Fine Grid of Prices

Notice that we still have to choose the frequency of the fine grid. We have used 1-minute as an example but this number is clearly also asset dependent. An asset with $N = 2{,}000$ new quotes on average per day should have a finer grid than an asset with $N = 50$ new quotes on average per day.

We ought to have at least one quote per interval on the fine grid. So we should definitely have that $m < N$. However, the distribution of quotes is typically very uneven throughout the day and so setting m close to N is likely to yield many intervals without new quotes. We can capture the distribution of quotes across time on each day by computing the standard deviation of $t(i+1) - t(i)$ across i on each day.

The total number of new quotes, N, will differ across days and so will the standard deviation of the quote time intervals. Looking at the descriptive statistics of N and the standard deviation of quote time intervals across days is likely to yield useful guidance on the choice of m.

5.3 Dealing with Data Gaps from Overnight Market Closures

In risk management we are typically interested in the volatility of 24-hour returns (the return from the close of day t to the close on day $t+1$) even if the market is only open, say, 8 hours per day. In Chapter 4 we estimated GARCH models on daily returns from

closing prices. The volatility forecasts from GARCH are therefore by construction 24-hour return volatilities.

If we care about 24-hour return volatility and we only have intraday returns from market open to market close, then the RV measure computed on intraday returns, call it RV_{t+1}^{Open}, must be adjusted for the return in the overnight gap from close on day t to open on day $t+1$. There are three ways to make this adjustment.

First, we can simply scale up the market-open RV measure using the unconditional variance estimated from daily squared returns:

$$RV_{t+1}^{24H} = \left(\frac{\sum_{t=1}^{T} R_t^2}{\sum_{t=1}^{T} RV_t^{Open}} \right) RV_{t+1}^{Open}$$

Second, we can add to RV_{t+1}^{Open} the squared return constructed from the close on day t to the open on day $t+1$:

$$RV_{t+1}^{24H} = \ln\left(S_{t+1}^{Open}/S_t^{Close}\right)^2 + RV_{t+1}^{Open}$$

Notice that this sum puts equal weight on the two terms and thus a relatively high weight on the close-to-open gap for which little information is available. Note also that S_t^{Close} is simply the daily price observation that we denoted S_t in the previous chapters.

A third, but more cumbersome approach is to find optimal weights for the two terms. This can be done by minimizing the variance of the RV_{t+1}^{24H} estimator subject to having a bias of zero.

When computing optimal weights typically a much larger weight is found for RV_{t+1}^{Open} than for $\ln\left(S_{t+1}^{Open}/S_t^{Close}\right)^2$. This suggests that scaling up the RV_{t+1}^{Open} may be the better of the two first approaches to correcting for the overnight gap.

5.4 Alternative RV Estimators Using Tick-by-Tick Data

There is an alternative set of RV estimators that avoid the construction of a time grid altogether and instead work directly with the irregularly spaced tick-by-tick data. Let the ith tick return on day $t+1$ be defined by

$$R_{t(i+1)} = \ln(S_{t(i+1)}) - \ln(S_{t(i)})$$

Then the tick-based RV estimator is defined by

$$RV_{t+1}^{Tick} = \sum_{i=1}^{N-1} R_{t(i+1)}^2$$

Notice that tick-time sampling avoids sampling the same observation multiple times, which could happen on a fixed grid if the grid is too fine compared with the number of available intraday prices.

The preceding simple tick-based RV estimator can be extended by allowing for autocorrelation in the tick-time returns.

The optimality of grid-based versus tick-based RV estimators depends on the structure of the market for the asset and on its liquidity. The majority of academic research relies on grid-based RV estimators.

5.5 Price and Quote Data Errors

The construction of the intraday price grid is perhaps the most challenging task when estimating and forecasting volatility using realized variance. The raw intraday price data contains observations randomly spaced in time and the sheer volume of data can be enormous when investigating many assets over long time periods.

The construction of the grid of prices is complicated by the presence of data errors. A data error is broadly defined as a quoted price that does not conform to the real situation of the market. Price data errors could take several forms:

- Decimal errors; for example, when a bid price changes from 1.598 to 1.603 but a 1.503 is reported instead of 1.603.
- Test quotes: These are quotes sent by a contributor at early mornings or at other inactive times to test the system. They can be difficult to catch since the prices may look plausible.
- Repeated ticks: These are sent automatically by contributors. If sent frequently, then they can obstruct the filtering of a few informative quotes sent by other contributors.
- Tick copying: Contributors automatically copy and resend quotes of other contributors to show a strong presence in the market. Sometimes random error is added so as to hide the copying aspect.
- Scaling problems: The scale of the price of an asset may differ by contributor and it may change over time without notice.

Given the size of intraday data sets it is impossible to manually check for errors. Automated filters must be developed to catch errors of the type just listed. The challenges of filtering intraday data has created a new business for data vendors. OlsenData.com and TickData.com are examples of data vendors that sell filtered as well as raw intraday data.

6 Range-Based Volatility Modeling

The construction of daily realized volatilities relies on the availability of intraday prices on relatively liquid assets. For markets that are not liquid, or for assets where

historical information on intraday prices is not available the intraday range presents a convenient alternative.

The intraday price range is based on the intraday high and intraday low price. Casual browsing of the web (see for example finance.yahoo.com) reveals that these intraday high and low prices are easily available for many assets far back in time. Range-based variance proxies are therefore easily computed.

6.1 Range-Based Proxies for Volatility

Let us define the range of the log prices to be

$$D_t = \ln(S_t^{High}) - \ln\left(S_t^{Low}\right) = \ln\left(S_t^{High}/S_t^{Low}\right)$$

where S_t^{High} and S_t^{Low} are the highest and lowest prices observed during day t.

We can show that if the log return on the asset is normally distributed with zero mean and variance, σ^2, then the expected value of the squared range is

$$E\left[D_t^2\right] = 4\ln(2)\sigma^2$$

A natural range-based estimate of volatility is therefore

$$\sigma^2 = \frac{1}{4\ln(2)}\frac{1}{T}\sum_{t=1}^{T} D_t^2$$

The range-based estimate of variance is simply a constant times the average squared range. The constant is $\frac{1}{4\ln(2)} \approx 0.361$.

The range-based estimate of unconditional variance suggests that a range proxy for the daily variance can be constructed as

$$RP_t = \tfrac{1}{4\ln(2)} D_t^2 \approx 0.361 D_t^2$$

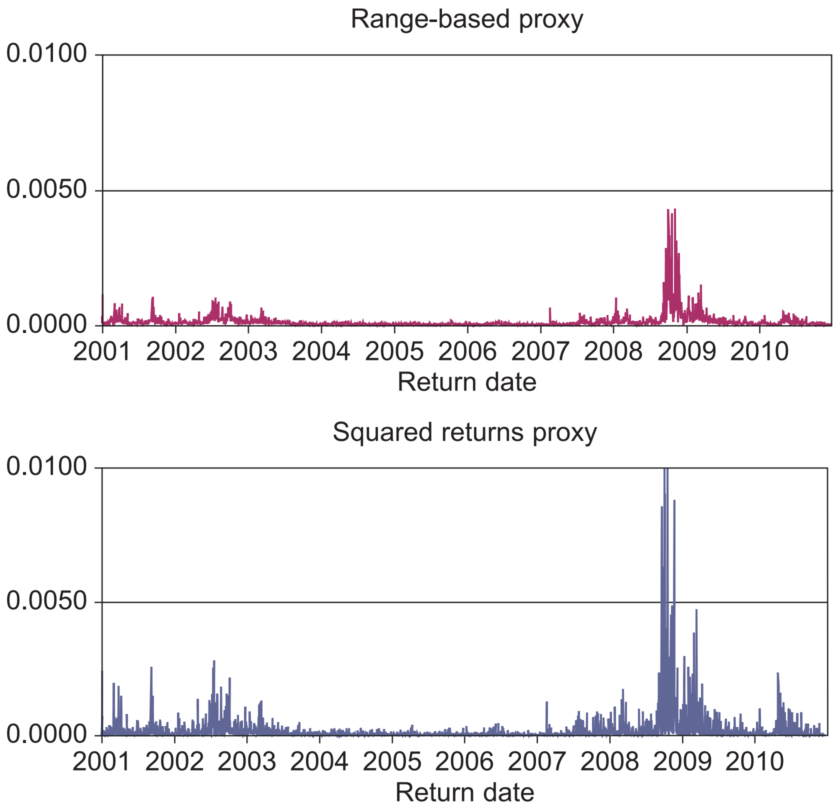
The top panel of Figure 5.6 plots RP_t for the S&P 500 data.

Notice how much less noisy the range is than the daily squared returns that are shown in the bottom panel.

Figure 5.7 shows the autocorrelation of RP_t in the top panel. The first-order autocorrelation in the range-based variance proxy is around 0.60 (top panel) whereas it is only half of that in the squared-return proxy (bottom panel). Furthermore, the range-based autocorrelations are much smoother and thus give a much more reliable picture of the persistence in variance than do the squared returns in the bottom panel.

This range-based volatility proxy does not make use of the daily open and close prices, which are also easily available and which also contain information about the 24-hour volatility. Assuming again that the asset log returns are normally distributed

Figure 5.6 Range-based variance proxy (top) and squared returns (bottom).



Notes: We use the daily range proxy for variance computed from the intraday high and low prices (top panel) and the daily close-to-close squared returns (bottom panel).

with zero mean and variance, σ^2, then a more accurate range-based proxy can be derived as

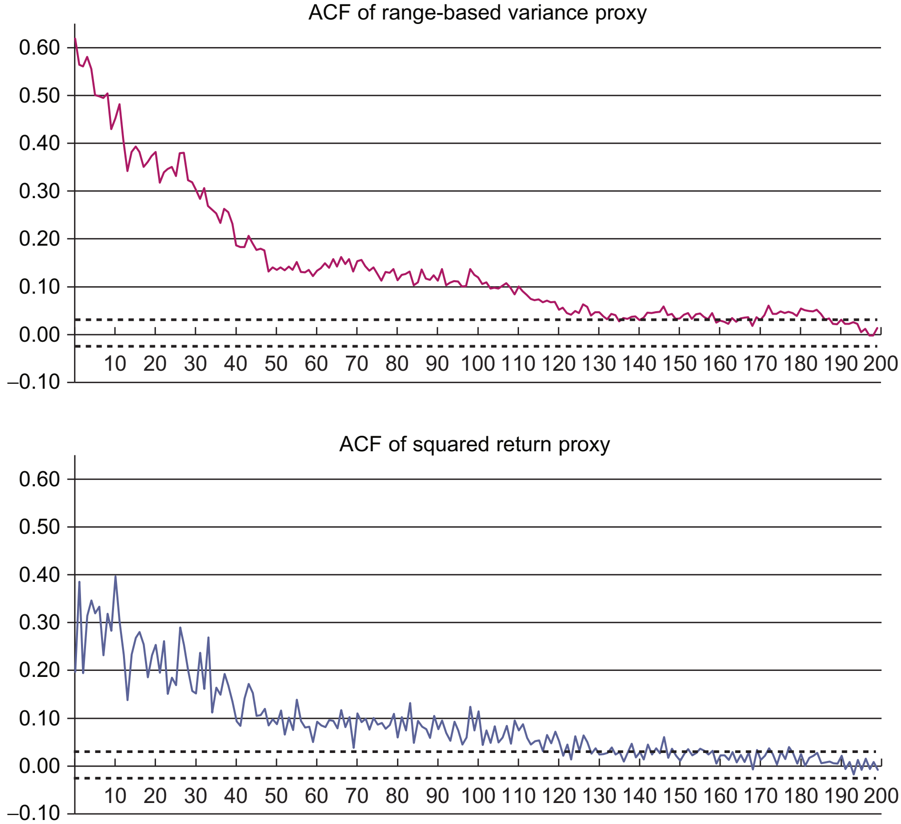
$$RP_t = \frac{1}{2}D_t^2 - (2\ln(2) - 1)\ln\left(S_t^{Close}/S_t^{Open}\right)^2$$

In the more general case where the mean return is not assumed to be zero the following range-based volatility proxy is available:

$$\begin{aligned} RP_t = {} & \ln\left(S_t^{High}/S_t^{Open}\right)\left[\ln\left(S_t^{High}/S_t^{Open}\right) - \ln\left(S_t^{Close}/S_t^{Open}\right)\right] \\ & + \ln\left(S_t^{Low}/S_t^{Open}\right)\left[\ln\left(S_t^{Low}/S_t^{Open}\right) - \ln\left(S_t^{Close}/S_t^{Open}\right)\right] \end{aligned}$$

All of these proxies are derived assuming that the true variance is constant, so that, for example, 30 days of high, low, open, and close information can be used to estimate the

Figure 5.7 Autocorrelation of the range-based variance proxy (top) and autocorrelation of squared returns (bottom) with Bartlett standard errors (dashed).



Notes: We compute autocorrelations from the daily range proxy for variance computed using the intraday high and low prices (top panel) and from the daily close-to-close squared returns (bottom panel) using the S&P 500 index.

(constant) volatility for that period. We instead want to use the range-based proxies as input into a dynamic forecasting model for volatility in line with the GARCH models in Chapter 4 and the HAR models in this chapter.

6.2 Forecasting Volatility Using the Range

Perhaps the simplest approach to using RP_t in a forecasting model is to use it in place of RV in the earlier AR and HAR models. Although RP_t may be more noisy than RV, the HAR approach should yield good forecasting results because the HAR model structure imposes a lot of smoothing.

Several studies have found that the log range is close to normally distributed as follows:

$$\ln(RP_t) \sim N\left(\mu_{RP}, \sigma_{RP}^2\right)$$

Recall that RV in logs is also close to normally distributed as well as we saw in Figure 5.3.

The strong persistence of the range as well as the log normal property suggest a log HAR model of the form

$$\ln(RP_{t+1}) = \phi_0 + \phi_D \ln(RP_{D,t}) + \phi_W \ln(RP_{W,t}) + \phi_M \ln(RP_{M,t}) + \varepsilon_{t+1}$$

where we have that

$$\begin{aligned} RP_{D,t} &\equiv RP_t \\ RP_{W,t} &\equiv \left[RP_{t-4,t} + RP_{t-3,t} + RP_{t-2,t} + RP_{t-1,t} + RP_t\right]/5 \\ RP_{M,t} &\equiv \left[RP_{t-20,t} + RP_{t-19,t} + \cdots + RP_t\right]/21 \end{aligned}$$

The range-based proxy can also be used as a regressor in GARCH-X models, for example

$$\begin{aligned} R_{t+1} &= \sigma_{t+1} z_{t+1}, \text{ where} \\ \sigma_{t+1}^2 &= \omega + \alpha R_t^2 + \beta \sigma_t^2 + \gamma RP_t \end{aligned}$$

A purely range-based model can be defined as

$$\begin{aligned} R_{t+1} &= \sigma_{t+1} z_{t+1}, \text{ where} \\ \sigma_{t+1}^2 &= \omega + \alpha RP_t + \beta \sigma_t^2 \end{aligned}$$

Finally, a Realized-GARCH style model (let us call it Range-GARCH) can be defined via

$$\begin{aligned} R_{t+1} &= \sigma_{t+1} z_{t+1}, \text{ where} \\ \sigma_{t+1}^2 &= \omega + \alpha R_t^2 + \beta \sigma_t^2 + \gamma RP_t, \text{ and} \\ RP_t &= \omega_{RP} + \beta_{RP} \sigma_t^2 + \varepsilon_t \end{aligned}$$

The Range-GARCH model can be estimated using bivariate maximum likelihood techniques using historical data on return, R_t, and on the range proxy, RP_t.

ES and *VaR* can be constructed in the RP-based models in the same way as in the RV-based models by assuming that z_{t+1} is i.i.d. normal where $z_{t+1} = R_{t+1}/\sigma_{t+1}$ in the GARCH-style models or $z_{t+1} = R_{t+1}/\sqrt{E_t(RP_{t+1})}$ in the HAR model.

6.3 Range-Based versus Realized Variance

There is convincing empirical evidence that for very liquid securities the RV modeling approach is useful for risk management purposes. The intuition is that using the intraday returns gives a very reliable estimate of today's variance, which in turn helps forecast tomorrow's variance. In standard GARCH models on the other hand, today's variance is implicitly calculated using exponentially declining weights on many past daily squared returns, where the exact weighting scheme depends on the estimated parameters. Thus the GARCH estimate of today's variance is heavily model dependent, whereas the realized variance for today is calculated exclusively from today's squared intraday returns. When forecasting the future, knowing where you are today is key. Unfortunately in variance forecasting, knowing where you are today is not a trivial matter since variance is not directly observable.

While the realized variance approach has clear advantages it also has certain shortcomings. First of all it clearly requires high-quality intraday returns to be feasible. Second, it is very easy to calculate daily realized volatilities from 5-minute returns, but it is not at all a trivial matter to construct at 10-year data set of 5-minute returns.

Figure 5.5 illustrates that the observed intraday price can be quite noisy compared with the fundamental but unobserved price. Therefore, realized variance measures based on intraday returns can be noisy as well. This is especially true for securities with wide bid–ask spreads and infrequent trading. Notice on the other hand that the range-based variance measure discussed earlier is relatively immune to the market microstructure noise. The true maximum can easily be calculated as the observed maximum less one half of the bid–ask spread, and the true minimum as the observed minimum plus one half of the bid–ask price. The range-based variance measure thus has clear advantages in less liquid markets.

In the absence of trading imperfections, however, range-based variance proxies can be shown to be only about as useful as 4-hour intraday returns. Furthermore, as we shall see in Chapter 7, the idea of realized variance extends directly to realized covariance and correlation, whereas the range-based covariance and correlation measures are less obvious.

7 GARCH Variance Forecast Evaluation Revisited

In the previous chapter we briefly introduced regressions using daily squared returns to evaluate the GARCH model forecasts. But we quickly argued that daily returns are too noisy to proxy for observed daily variance. In this chapter we have developed more informative proxies based on RV and RP and they should clearly be useful for variance forecast evaluation.

The realized variance measure can be used instead of the squared return for evaluating the forecasts from variance models. If only squared returns are available then

we can run the regression

$$R_{t+1}^2 = b_0 + b_1 \sigma_{t+1|t}^2 + \varepsilon_{t+1}$$

where $\sigma_{t+1|t}^2$ is the forecast from the GARCH model.

If we have RV-based estimates available then we would instead run the regression

$$RV_{t+1}^{Avr} = b_0 + b_1 \sigma_{t+1|t}^2 + \varepsilon_{t+1}$$

where we have used the Average RV estimator as an example.

The range-based proxy could of course also be used instead of the squared return for evaluating the forecasts from variance models. Thus we could run the regression

$$RP_{t+1} = b_0 + b_1 \sigma_{t+1|t}^2 + \varepsilon_{t+1}$$

where RP_{t+1} can be constructed for example using

$$RP_{t+1} = 0.361 \ln \left(S_{t+1}^{High} / S_{t+1}^{Low} \right)^2$$

Using R_{t+1}^2 on the left-hand side of these regressions is likely to yield the finding that the volatility forecast is poor. The fit of the regression will be low but notice that this does not necessarily mean that the volatility forecast is poor. It could also mean that the volatility proxy is poor. If regressions using RV_{t+1}^{Avr} or RP_{t+1} yield a much better fit than the regression using R_{t+1}^2 then the volatility forecast is much better than suggested by the noisy squared-return proxy.

8 Summary

Realized volatility and range-based volatility are likely to be much more informative about daily volatility than is the daily squared return. This fact has important implications for the evaluation of volatility forecasts but it has even more important implications for volatility forecast construction. If intraday information is available then it should be used to construct more accurate volatility forecasts than those that can be constructed from daily returns alone. This chapter has introduced a number of practical approaches to volatility forecasting using intraday information.

Further Resources

The classic references on realized volatility include Andersen et al. (2001, 2003), and Barndorff-Nielsen and Shephard (2002). See the survey in Andersen et al. (2010) for a thorough literature review.

The HAR model for RV was developed in Corsi (2009) and has been used in Andersen et al. (2007b) among others. Engle (2002) suggested RV in the GARCH-X model

and the Realized GARCH model was developed in Hansen et al. (2011). See also the HEAVY model in Shephard and Sheppard (2010).

The crucial impact on RV of liquidity and market microstructure effects more generally has been investigated in Andersen et al. (2011), Bandi and Russell (2006), and Ait-Sahalia and Mancini (2008).

The choice of sampling frequency has been analyzed by Ait-Sahalia et al. (2005), and Bandi and Russell (2008). The volatility signature plot was suggested in Andersen et al. (1999). The Average RV estimator is discussed in Zhang et al. (2005). The RV estimates corrected for return autocorrelations were developed by Zhou (1996), Barndorff-Nielsen et al. (2008), and Hansen and Lunde (2006).

The use of RV in volatility forecast evaluation was pioneered by Andersen and Bollerslev (1998). See also Andersen et al. (2004, 2005) and Patton (2011).

The use of RV in risk management is discussed in Andersen et al. (2007a), and the use of RV in portfolio allocation is developed in Bandi et al. (2008) and Fleming et al. (2003).

For forecasting applications of RV see Martens (2002), Thomakos and Wang (2003), Pong et al. (2004), Koopman et al. (2005), and Maheu and McCurdy (2011).

For treating overnight gaps see Hansen and Lunde (2005), and for data issues in RV construction see Brownlees and Gallo (2006), Muller (2001), and Dacorogna et al. (2001).

Range-based estimates variance models are introduced in Parkinson (1980) and Garman and Klass (1980), and more recent contributions include Rogers and Satchell (1991) and Yang and Zhang (2000). Range-based models of dynamic variance are developed in Azalideh et al. (2002), Brandt and Jones (2006), and Chou (2005), and they are surveyed in Chou et al. (2009). Brandt and Jones (2006) use the range rather than the squared return as the fundamental innovation in an EGARCH model and find that the range improves the model's variance forecasts significantly.

References

Ait-Sahalia, Y., Mancini, L., 2008. Out of sample forecasts of quadratic variation. J. Econom. 147, 17–33.

Ait-Sahalia, Y., Mykland, P.A., Zhang, L., 2005. How often to sample a continuous-time process in the presence of market microstructure noise. Rev. Financ. Stud. 18, 351–416.

Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. Int. Econ. Rev. 39, 885–905.

Andersen, T., Bollerslev, T., Christoffersen, P., Diebold, F.X., 2007a. Practical volatility and correlation modeling for financial market risk management. In: Carey, M., Stulz, R. (Eds.), The NBER Volume on Risks of Financial Institutions. University of Chicago Press, Chicago, IL.

Andersen, T.G., Bollerslev, T., Diebold, F.X., 2007b. Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility. Rev. Econ. Stat. 89, 701–720.

Andersen, T.G., Bollerslev, T., Diebold, F.X., 2010. Parametric and nonparametric measurements of volatility. In: Ait-Sahalia, Y., Hansen, L.P. (Eds.), Handbook of Financial Econometrics. North-Holland, Amsterdam, The Netherlands, pp. 67–138.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 1999. (Understanding, optimizing, using and forecasting) Realized volatility and correlation, published in revised form as “Great Realizations.” Risk March 2000, 105–108.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001. The distribution of exchange rate volatility. J. Am. Stat. Assoc. 96, 42–55.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579–625.

Andersen, T.G., Bollerslev, T., Meddahi, N., 2004. Analytic evaluation of volatility forecasts. Int. Econ. Rev. 45, 1079–1110.

Andersen, T.G., Bollerslev, T., Meddahi, N., 2005. Correcting the errors: Volatility forecast evaluation using high-frequency data and realized volatilities. Econometrica 73, 279–296.

Andersen, T., Bollerslev, T., Meddahi, N., 2011. Realized volatility forecasting and market microstructure noise. J. Econom. 160, 220–234.

Azalideh, S., Brandt, M., Diebold, F.X., 2002. Range-based estimation of stochastic volatility models. J. Finance 57, 1047–1091.

Bandi, F., Russell, J., 2006. Separating microstructure noise from volatility. J. Financ. Econ. 79, 655–692.

Bandi, F., Russell, J., 2008. Microstructure noise, realized volatility, and optimal sampling. Rev. Econ. Stud. 75, 339–369.

Bandi, F., Russell, J., Zhu, Y., 2008. Using high-frequency data in dynamic portfolio choice. Econom. Rev. 27, 163–198.

Barndorff-Nielsen, O.E., Hansen, P., Lunde, A., Shephard, N., 2008. Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. Econometrica 76, 1481–1536.

Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realised volatility and its use in estimating stochastic volatility models. J. Royal Stat. Soc. B 64, 253–280.

Brandt, M., Jones, C., 2006. Volatility forecasting with range-based EGARCH models. J. Bus. Econ. Stat. 24, 470–486.

Brownlees, C., Gallo, G.M., 2006. Financial econometric analysis at ultra-high frequency: Data handling concerns. Comput. Stat. Data Anal. 51, 2232–2245.

Chou, R., 2005. Forecasting financial volatilities with extreme values: The conditional autoregressive range. J. Money Credit Bank. 37, 561–82.

Chou, R., Chou, H.-C., Liu, N., 2009. Range volatility models and their applications in finance. In: Lee, C.-F., Lee, J. (Eds.), Handbook of Quantitative Finance and Risk Management. Springer, New York, NY, pp. 1273–1281.

Corsi, F., 2009. A simple approximate long memory model of realized volatility. J. Financ. Econom. 7, 174–196.

Dacorogna, M., Gencay, R., Muller, U., Olsen, R., Pictet, O., 2001. An Introduction to High-Frequency Finance. Academic Press, San Diego, CA.

Engle, R., 2002. New frontiers in ARCH models. J. Appl. Econom. 17, 425–446.

Fleming, J., Kirby, C., Oestdiek, B., 2003. The economic value of volatility timing using ‘realized’ volatility, with Jeff Fleming and Chris Kirby. J. Financ. Econ. 67, 473–509.

Garman, M., Klass, M., 1980. On the estimation of securities price volatilities from historical data. J. Bus. 53, 67–78.

Hansen, P., Huang, Z., Shek, H., 2011. Realized GARCH: A joint model for returns and realized measures of volatility. J. Appl. Econom. forthcoming.

Hansen. P., Lunde, A., 2005. A realized variance for the whole day based on intermittent high-frequency data. J. Financ. Econom. 3, 525–554.
Hansen, P., Lunde, A., 2006. Realized variance and market microstructure noise. J. Bus. Econ. Stat. 24, 127–161.
Koopman, S.J., Jungbacker, B., Hol, E., 2005. Forecasting daily variability of the S&P 100 stock index using historical, realized and implied volatility measures. J. Empir. Finance 12, 445–475.
Maheu, J., McCurdy, T., 2011. Do high-frequency measures of volatility improve forecasts of return distributions? J. Econom. 160, 69–76.
Martens, M., 2002. Measuring and forecasting S&P 500 index futures volatility using high-frequency data. J. Futures Mark. 22, 497–518.
Muller, U., 2001. The Olsen filter for data in finance. Working paper, O&A Research Group. Available from: http://www.olsendata.com.
Parkinson, M., 1980. The extreme value method for estimating the variance of the rate of return. J. Bus. 53, 61–65.
Patton, A., 2011. Volatility forecast comparison using imperfect volatility proxies. J. Econom. 160, 246–256.
Pong, S., Shackleton, M.B., Taylor, S.J., Xu, X., 2004. Forecasting currency volatility: A comparison of implied volatilities and AR(FI)MA models. J. Bank. Finance 28, 2541–2563.
Rogers, L., Satchell, S., 1991. Estimating variance from high, low and closing prices. Ann. Appl. Probab. 1, 504–512.
Shephard, N., Sheppard, K., 2010. Realizing the future: Forecasting with high-frequency-based volatility (HEAVY) models. J. Appl. Econom. 25, 197–231.
Thomakos, D.D., Wang, T., 2003. Realized volatility in the futures market. J. Empir. Finance 10, 321–353.
Yang, D., Zhang, Q., 2000. Drift-independent volatility estimation based on high, low, open, and close prices. J. Bus. 73, 477–491.
Zhang, L., Mykland, P.A., Ait-Sahalia, Y., 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. J. Am. Stat. Assoc. 100, 1394–1411.
Zhou, B., 1996. High-frequency data and volatility in foreign exchange rates. J. Bus. Econ. Stat. 14, 45–52.

Empirical Exercises

Open the Chapter5Data.xlsx file from the web site.

1. Run a regression of daily squared returns on the variance forecast from the GARCH model with a leverage term from Chapter 4. Include a constant term in the regression

$$R_{t+1}^2 = b_0 + b_1\sigma_{t+1}^2 + e_{t+1}$$

(*Excel hint*: Use the function LINEST.) What is the fit of the regression as measured by the R^2? Is the constant term significantly different from zero? Is the coefficient on the forecast significantly different from one?

2. Run a regression using *RP* instead of the squared returns as proxies for observed variance; that is, regress

$$RP_{t+1} = b_0 + b_1\sigma^2_{t+1} + e_{t+1}, \text{ where}$$
$$RP_{t+1} = \frac{1}{4\ln(2)}D^2_{t+1}$$

Is the constant term significantly different from zero? Is the coefficient on the forecast significantly different from one? What is the fit of the regression as measured by the R^2? Compare your answer with the R^2 from exercise 1.

3. Run a regression using *RV* instead of the squared returns as proxies for observed variance; that is, regress

$$RV_{t+1} = b_0 + b_1\sigma^2_{t+1} + e_{t+1}$$

Is the constant term significantly different from zero? Is the coefficient on the forecast significantly different from one? What is the fit of the regression as measured by the R^2? Compare your answer with the R^2 from exercises 1 and 2.

4. Estimate a HAR model in logarithms on the RP data you constructed in exercise 2. Use the next day's RP on the left-hand side and use daily, weekly, and monthly regressors on the right-hand side. Compute the regression fit.
5. Estimate a HAR model in logarithms on the RV data. Use the next day's RV on the left-hand side and use daily, weekly, and monthly regressors on the right-hand side. Compare the regression fit from this equation with that from exercise 4.

The answers to these exercises can be found in the Chapter5Results.xls file, which can be found on the companion site.

For more information see the companion site at
http://www.elsevierdirect.com/companions/9780123744487

6 Nonnormal Distributions

1 Chapter Overview

We now turn to the final part of the stepwise univariate distribution modeling approach, namely accounting for conditional nonnormality in portfolio returns. In Chapter 1, we saw that asset returns are not normally distributed. If we construct a simple histogram of past returns on the S&P 500 index, then it will not conform to the density of the normal distribution: The tails of the histogram are fatter than the tails of the normal distribution, and the histogram is more peaked around zero. From a risk management perspective, the fat tails, which are driven by relatively few but very extreme observations, are of most interest. These extreme observations can be symptoms of liquidity risk or event risk as defined in Chapter 1.

One motivation for the time-varying variance models discussed in Chapters 4 and 5 is that they are capable of accounting for some of the nonnormality in the daily returns. For example a GARCH(1,1) model with normally distributed shocks, $z_t = R_t/\sigma_t$ will imply a nonnormal distribution of returns R_t because the distribution of returns is a function of all the past return variances σ_i^2, $i = 1, 2, \ldots, t$.

GARCH models with normal shocks by definition do not capture what we call conditional nonnormality in the returns. Returns are conditionally normal if the shocks z_t are normally distributed. Histograms from shocks, (i.e. standardized returns) typically do not conform to the normal density. Figure 6.1 illustrates this point.

The top panel shows the histogram of the raw returns superimposed on the normal distribution and the bottom panel shows the histogram of the standardized returns superimposed on the normal distribution as well. The volatility model used to standardize the returns is the NGARCH(1,1) model, which includes a leverage effect. Notice that while the bottom histogram conforms more closely to the normal distribution than does the top histogram, there are still some systematic deviations, including fat tails and a more pronounced peak around zero.

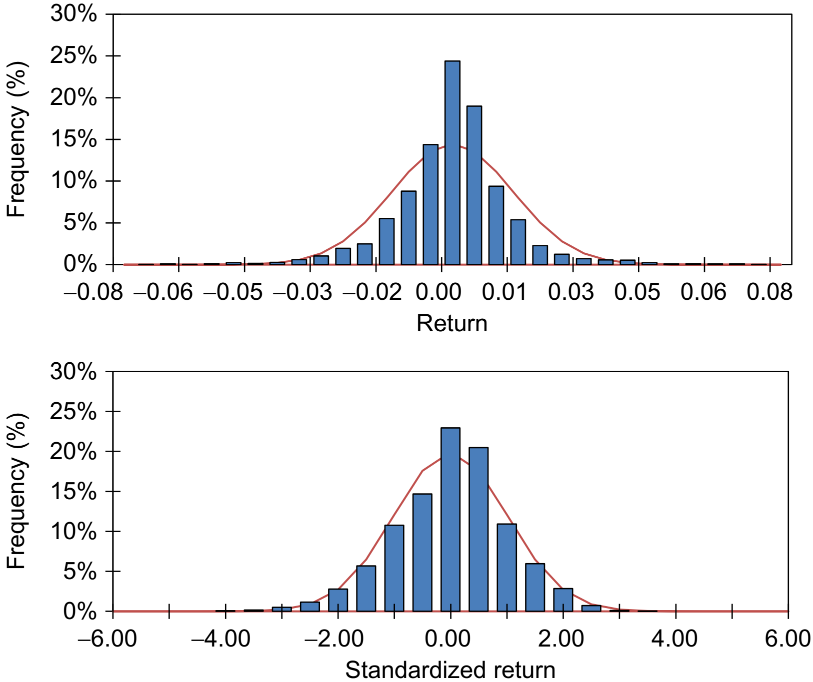
2 Learning Objectives

We will analyze the conditional nonnormality in several ways:

1. We introduce the quantile-quantile (QQ) plot, which is a graphical tool better at describing tails of distributions than the histogram.

Elements of Financial Risk Management. DOI: 10.1016/B978-0-12-374448-7.00006-3

Figure 6.1 Histogram of daily S&P 500 returns (top panel) and histogram of GARCH shocks (bottom panel).



Notes: The top panel shows a histogram of daily S&P 500 returns and the bottom panel shows a histogram of returns standardized by the dynamic variance from a GARCH model.

2. We define the Filtered Historical Simulation approach, which combines GARCH with Historical Simulation.
3. We introduce the simple Cornish-Fisher approximation to *VaR* in nonnormal distributions.
4. We consider the standardized Student's t distribution and discuss the estimation of it.
5. We extend the Student's t distribution to a more flexible asymmetric version.
6. We consider extreme value theory for modeling the tail of the conditional distribution.

For each of these methods we will provide the Value-at-Risk and the expected shortfall formulas.

Throughout this chapter, we will assume that we are working with a time series of portfolio returns using today's portfolio weights and past returns on the underlying assets in the portfolio. Therefore, we are modeling a univariate time series. We will assume that the portfolio variance has already been modeled using the methods presented in Chapters 4 and 5.

Working with the univariate time series of portfolio returns is convenient from a modeling perspective but it has the disadvantage of being conditional on exactly the current set of portfolio weights. If the weights are changed, then the portfolio distribution modeling will have to be redone. Multivariate risk models will be studied in Chapters 7–9.

3 Visualizing Nonnormality Using QQ Plots

As in Chapter 2, consider a portfolio of n assets. If we today own $N_{i,t}$ units or shares of asset i then the value of the portfolio today is

$$V_{PF,t} = \sum_{i=1}^{n} N_{i,t} S_{i,t}$$

Using today's portfolio holdings but historical asset prices we can compute the history of (pseudo) portfolio values. For example, yesterday's portfolio value is

$$V_{PF,t-1} = \sum_{i=1}^{n} N_{i,t} S_{i,t-1}$$

The log return can now be defined as

$$R_{PF,t} = \ln\left(V_{PF,t}/V_{PF,t-1}\right)$$

Allowing for a dynamic variance model we can write

$$R_{PF,t} = \sigma_{PF,t} z_t, \quad \text{with } z_t \overset{i.i.d.}{\sim} D(0,1)$$

where $\sigma_{PF,t}$ is the conditional volatility forecast constructed using the methods in the previous two chapters.

The focus in this chapter is on modeling the distribution of the innovations, $D(0,1)$, which has a mean of zero and a standard deviation of 1. So far, we have relied on setting $D(0,1)$ to $N(0,1)$, but we now want to assess the problems of the normality assumption in risk management, and we want to suggest viable alternatives.

Before we venture into the particular formulas for suitable nonnormal distributions, let us first introduce a valuable visual tool for assessing nonnormality, which we will also use later as a diagnostic check on nonnormal alternatives. The tool is commonly known as a quantile-quantile (QQ) plot, and the idea is to plot the empirical quantiles of the calculated returns, which is simply the returns ordered by size, against the corresponding quantiles of the normal distribution. If the returns are truly normal, then the graph should look like a straight line at a 45-degree angle. Systematic deviations from the 45-degree line signal that the returns are not well described by the normal

distribution. QQ plots are, of course, particularly relevant to risk managers who care about Value-at-Risk, which itself is a quantile.

The QQ plot is constructed as follows: First, sort all standardized returns $z_t = R_{PF,t}/\sigma_{PF,t}$ in ascending order, and call the ith sorted value z_i. Second, calculate the empirical probability of getting a value below the actual as $(i-0.5)/T$, where T is the total number of observations. The subtraction of 0.5 is an adjustment for using a continuous distribution on discrete data.

Calculate the standard normal quantiles as $\Phi^{-1}_{(i-0.5)/T}$, where Φ^{-1} denotes the inverse of the standard normal density as before. We can then scatter plot the standardized and sorted returns on the Y-axis against the standard normal quantiles on the X-axis as follows:

$$\{X_i, Y_i\} = \left\{\Phi^{-1}_{(i-0.5)/T}, z_i\right\}$$

If the data were normally distributed, then the scatterplot should conform roughly to the 45-degree line.

Figure 6.2 shows a QQ plot of the daily S&P 500 returns from Chapter 1. The top panel uses standardized returns from the unconditional standard deviation, σ_{PF}, so that $z_t = R_{PF,t}/\sigma_{PF}$, and the bottom panel uses returns standardized by an NGARCH(1,1) with a leverage effect, $z_t = R_{PF,t}/\sigma_{PF,t}$.

Notice that the GARCH model does capture some of the nonnormality in the returns, but some still remains. The patterns of deviations from the 45-degree line indicate that large positive returns are captured remarkably well by the normal GARCH model but that the model does not allow for a sufficiently fat left tail as compared with the data.

4 The Filtered Historical Simulation Approach

The Filtered Historical Simulation approach (FHS), which we present next, attempts to combine the best of the model-based with the best of the model-free approaches in a very intuitive fashion. FHS combines model-based methods of dynamic variance, such as GARCH, with model-free methods of distribution in the following way.

Assume we have estimated a GARCH-type model of our portfolio variance. Although we are comfortable with our variance model, we are not comfortable making a specific distributional assumption about the standardized returns, such as a normal distribution. Instead, we would like the past returns data to tell us about the distribution directly without making further assumptions.

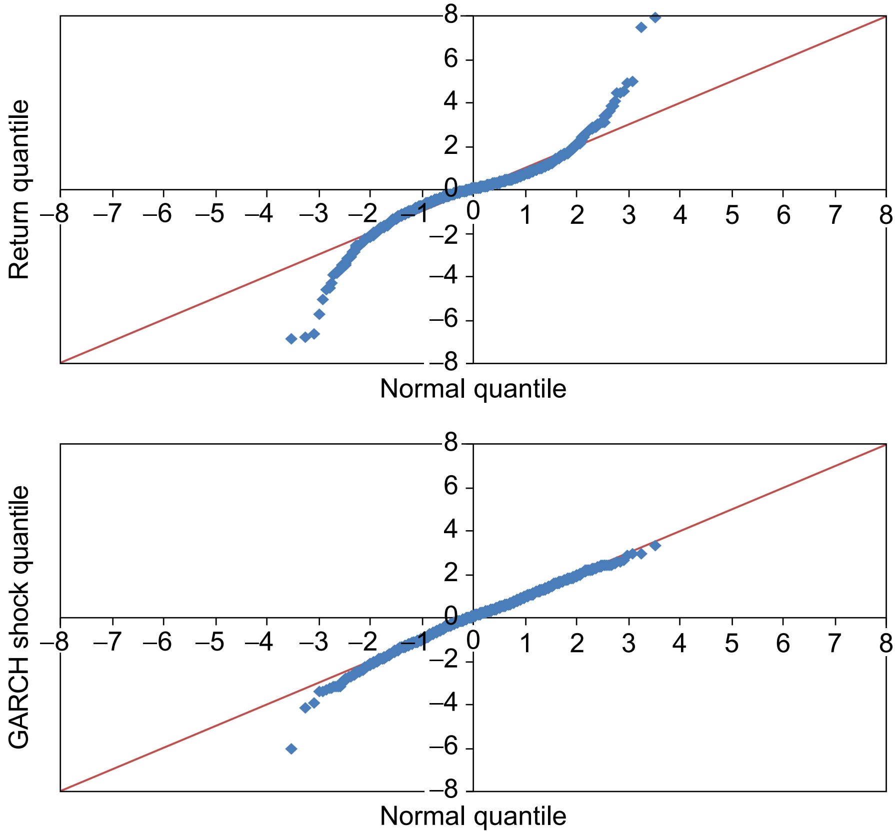
To fix ideas, consider again the simple example of a GARCH(1,1) model:

$$R_{PF,t+1} = \sigma_{PF,t+1} z_{t+1}$$

where

$$\sigma^2_{PF,t+1} = \omega + \alpha R^2_{PF,t} + \beta\sigma^2_{PF,t}$$

Figure 6.2 QQ plot of daily S&P 500 returns and GARCH shocks.



Notes: In the top panel we scatter plot the empirical quantiles of the S&P 500 returns (in standard deviations) against the normal distribution. In the bottom panel we scatter plot the empirical quantiles of the S&P 500 GARCH shocks against the quantiles of the normal distribution. The two red lines have a slope of one.

Given a sequence of past returns, $\{R_{PF,t+1-\tau}\}_{\tau=1}^{m}$, we can estimate the GARCH model and calculate past standardized returns from the observed returns and from the estimated standard deviations as

$$\hat{z}_{t+1-\tau} = R_{PF,t+1-\tau}/\sigma_{PF,t+1-\tau}, \quad \text{for } \tau = 1, 2, \ldots, m$$

We will refer to the set of standardized returns as $\{\hat{z}_{t+1-\tau}\}_{\tau=1}^{m}$.

We can simply calculate the 1-day *VaR* using the percentile of the database of standardized residuals as in

$$VaR_{t+1}^{p} = -\sigma_{PF,t+1} Percentile\left\{\{\hat{z}_{t+1-\tau}\}_{\tau=1}^{m}, 100p\right\}$$

At the end of Chapter 2, we introduced expected shortfall (*ES*) as an alternative risk measure to *VaR*. *ES* is defined as the expected return given that the return falls below the *VaR*. For the 1-day horizon, we have

$$ES^p_{t+1} = -E_t\left[R_{PF,t+1}|R_{PF,t+1} < -VaR^p_{t+1}\right]$$

The *ES* measure can be calculated from the historical shocks via

$$ES^p_{t+1} = -\frac{\sigma_{PF,t+1}}{p \cdot m}\sum_{i=1}^{m}\hat{z}_{t+1-\tau}\cdot \mathbf{1}\left(\hat{z}_{t+1-\tau} < -Percentile\left\{\left\{\hat{z}_{t+1-\tau}\right\}_{\tau=1}^{m}, 100p\right\}\right)$$

where the indicator function $\mathbf{1}(\bullet)$ returns a 1 if the argument is true and zero if not.

An interesting and useful feature of FHS as compared with the simple Historical Simulation approach introduced in Chapter 2 is that it can generate large losses in the forecast period, even without having observed a large loss in the recorded past returns. Consider the case where we have a relatively large negative z in our database, which occurred on a relatively low variance day. If this z gets combined with a high variance day in the simulation period then the resulting hypothetical loss will be large.

We close this section by reemphasizing that the FHS method suggested here combines a conditional model for variance with a Historical Simulation method for the standardized returns. FHS thus retains the key conditionality feature through σ_{t+1} but saves us from having to make assumptions beyond that the sample of historical zs provides a good description of the distribution of future zs. Note that this is very different from the standard Historical Simulation approach in which the sample of historical Rs is assumed to provide a good description of the distribution of future Rs.

5 The Cornish-Fisher Approximation to *VaR*

Filtered Historical Simulation offers a nice model-free approach to the conditional distribution. But FHS relies heavily on the recent series of observed shocks, z_t. If these shocks are interesting from a risk perspective (that is, they contain sufficiently many large negative values) then the FHS will deliver accurate results; if not, FHS may suffer.

We now consider a simple alternative way of calculating *VaR*, which has certain advantages. First, it does allow for skewness as well as excess kurtosis. Second, it is easily calculated from the empirical skewness and excess kurtosis estimates from the standardized returns, z_t. Third, it can be viewed as an approximation to the *VaR* from a wide range of conditionally nonnormal distributions.

We again start by defining standardized portfolio returns by

$$z_{t+1} = R_{PF,t+1}/\sigma_{PF,t+1} \overset{i.i.d.}{\sim} D(0, 1)$$

where $D(0, 1)$ denotes a distribution with a mean equal to 0 and a variance equal to 1. As in Chapter 4, *i.i.d.* denotes independently and identically distributed.

The Cornish-Fisher *VaR* with coverage rate p can then be calculated as

$$VaR^p_{t+1} = -\sigma_{PF,t+1} CF^{-1}_p$$

where

$$CF^{-1}_p = \Phi^{-1}_p + \frac{\zeta_1}{6}\left[(\Phi^{-1}_p)^2 - 1\right] + \frac{\zeta_2}{24}\left[(\Phi^{-1}_p)^3 - 3\Phi^{-1}_p\right] \\ - \frac{\zeta_1^2}{36}\left[2(\Phi^{-1}_p)^3 - 5\Phi^{-1}_p\right]$$

where ζ_1 is the skewness and ζ_2 is the excess kurtosis of the standardized returns, z_t. The Cornish-Fisher quantile can be viewed as a Taylor expansion around the normal distribution. Notice that if we have neither skewness nor excess kurtosis so that $\zeta_1 = \zeta_2 = 0$, then we simply get the quantile of the normal distribution

$$CF^{-1}_p = \Phi^{-1}_p, \quad \text{for } \zeta_1 = \zeta_2 = 0$$

Consider now for example the 1% *VaR*, where $\Phi^{-1}_{.01} \approx -2.33$. Allowing for skewness and kurtosis we can calculate the Cornish-Fisher 1% quantile as

$$CF^{-1}_p \approx -2.33 + 0.74\zeta_1 - 0.24\zeta_2 + 0.38\zeta_1^2$$

and the portfolio *VaR* can be calculated as

$$VaR^{.01}_{t+1} = -(-2.33 + 0.74\zeta_1 - 0.24\zeta_2 + 0.38\zeta_1^2)\sigma_{PF,t+1}$$

Thus, for example, if skewness equals –1 and excess kurtosis equals 4, then we get

$$VaR^{.01}_{t+1} = -(-2.33 - 0.74 - 0.24 \cdot 4 + 0.38)\sigma_{PF,t+1} = 3.63\sigma_{PF,t+1}$$

which is much higher than the *VaR* number from a normal distribution, which equals $2.33\sigma_{PF,t+1}$.

The expected shortfall can be derived as

$$ES^p_{t+1} = -\sigma_{PF,t+1} ES_{CF}(p)$$

where

$$ES_{CF}(p) = \frac{-\phi\left(CF^{-1}_p\right)}{p}\left[1 + \frac{\zeta_1}{6}\left(CF^{-1}_p\right)^3 + \frac{\zeta_2}{24}\left[\left(CF^{-1}_p\right)^4 - 2\left(CF^{-1}_p\right)^2 - 1\right]\right]$$

This derivation can be found in Appendix B. Recall from Chapter 2 that the *ES* for the normal distribution is

$$ES^p_{t+1} = \sigma_{PF,t+1}\frac{\phi\left(CF^{-1}_p\right)}{p}$$

which is also a special case of $ES_{CF}(p)$ when $\zeta_1 = \zeta_2 = 0$.

The CF approach is easy to implement and we avoid having to make an assumption about exactly which distribution fits the data best. However, exact distributions have advantages too. Perhaps most importantly for risk management, exact distributions allow us to compute *VaR* and *ES* for extreme probabilities (as we did in Chapter 2) for which the approximative CF may not be well-defined. Exact distributions also enable Monte Carlo simulation, which we will discuss in Chapter 8. We therefore consider useful examples of exact distributions next.

6 The Standardized *t* Distribution

Perhaps the most important deviations from normality we have seen are the fatter tails and the more pronounced peak in the distribution of z_t as compared with the normal distribution. The Student's t distribution captures these features. It is defined by

$$f_{t(d)}(x;d) = \frac{\Gamma((d+1)/2)}{\Gamma(d/2)\sqrt{d\pi}}(1+x^2/d)^{-(1+d)/2}, \quad \text{for } d > 0$$

The $\Gamma(\bullet)$ notation refers to the gamma function, which can be found in most quantitative software packages. Conveniently, the distribution has only one parameter, namely d. In the Student's t distribution we have the following first two moments:

$$\begin{aligned} E[x] &= 0, \quad \text{when } d > 1 \\ Var[x] &= d/(d-2) \quad \text{when } d > 2 \end{aligned}$$

We have already modeled variance using GARCH and other models and so we are interested in a distribution that has a variance equal to 1. The standardized t distribution—call it the $\tilde{t}(d)$ distribution—is derived from the Student's t to achieve this goal.

Define z by standardizing x so that

$$z = \frac{x - E[x]}{\sqrt{Var[x]}} = \frac{x}{\sqrt{d/(d-2)}}$$

The standardized $\tilde{t}(d)$ density is then defined by

$$f_{\tilde{t}(d)}(z;d) = C(d)\,(1+z^2/(d-2))^{-(1+d)/2}, \quad \text{for } d > 2$$

where

$$C(d) = \frac{\Gamma((d+1)/2)}{\Gamma(d/2)\sqrt{\pi(d-2)}}$$

Note that the standardized t distribution is defined so that the random variable z has mean equal to zero and a variance (and standard deviation) equal to 1. Note also that

the parameter d must be larger than two for the standardized distribution to be well defined.

The key feature of the $\tilde{t}(d)$ distribution is that the random variable, z, is taken to a power, rather than an exponential, which is the case in the standard normal distribution where

$$f(z) = (2\pi)^{-1/2} \exp\left(-z^2/2\right)$$

The power function driven by d will allow for the $\tilde{t}(d)$ distribution to have fatter tails than the normal; that is, higher values of $f_{\tilde{t}(d)}(\bullet)$ when z is far from zero.

The $\tilde{t}(d)$ distribution is symmetric around zero, and the mean (μ), variance $\left(\sigma^2\right)$, skewness (ζ_1), and excess kurtosis (ζ_2) of the distribution are

$$\begin{aligned}
\mu &\equiv E[z] = 0 \\
\sigma^2 &\equiv E[(z - E[z])^2] = 1 \\
\zeta_1 &\equiv E[z^3]/\sigma^3 = 0 \\
\zeta_2 &\equiv E[z^4]/\sigma^4 - 3 = 6/(d-4)
\end{aligned}$$

Thus, notice that d must be higher than 4 for the kurtosis to be well defined. Notice also that for large values of d the distribution will have an excess kurtosis of zero, and we can show that it converges to the standard normal distribution as d goes to infinity. Indeed, for values of d above 50, the $\tilde{t}(d)$ distribution is difficult to distinguish from the standard normal distribution.

6.1 Maximum Likelihood Estimation

Combining a dynamic volatility model such as GARCH with the standardized t distribution we can now specify our model portfolio returns as

$$R_{PF,t} = \sigma_{PF,t} z_t, \quad \text{with } z_t \overset{i.i.d.}{\sim} \tilde{t}(d)$$

If we ignore the fact that variance is estimated with error, we can treat the standardized return as a regular random variable, calculated as $z_t = R_{PF,t}/\sigma_{PF,t}$. The d parameter can then be estimated using maximum likelihood by choosing the d, which maximizes

$$\begin{aligned}
\ln L_1 &= \sum_{t=1}^{T} \ln(f_{\tilde{t}(d)}(z_t; d)) \\
&= T\{\ln(\Gamma((d+1)/2)) - \ln(\Gamma(d/2)) - \ln(\pi)/2 - \ln(d-2)/2\} \\
&\quad - \frac{1}{2}\sum_{t=1}^{T}(1+d)\ln(1 + (R_{PF,t}/\sigma_{PF,t})^2/(d-2))
\end{aligned}$$

Given that we have already modeled and estimated the portfolio variance $\sigma^2_{PF,t}$, and taken it as given, we can maximize $\ln L_1$ with respect to the parameter d only. This approach builds again on the quasi-maximum likelihood idea, and it is helpful in that we are only estimating few parameters at a time, in this case only one. The simplicity is important because we are relying on numerical optimization to estimate the parameters.

If we instead want to estimate the variance parameters and the d parameter simultaneously, we must adjust the distribution to take into account the variance, $\sigma^2_{PF,t}$, and we get

$$f(R_{PF,t}; d) = \frac{C(d)}{\sigma_{PF,t}}(1 + (R_{PF,t}/\sigma_{PF,t})^2/(d-2))^{-(1+d)/2}$$

To estimate all the parameters together, we must maximize the log-likelihood of the sample of returns, which can be written

$$\ln L_2 = \sum_{t=1}^{T} \ln(f(R_{PF,t}; d)) = \ln L_1 - \sum_{t=1}^{T} \ln(\sigma^2_{PF,t})/2$$

When we maximize $\ln L_2$ over all the parameters simultaneously, including the GARCH parameters implicit in $\sigma^2_{PF,t}$, then we will typically get more precise parameter estimates compared with stepwise estimation of the GARCH parameters first and the distribution parameters second.

As a simple univariate example of the difference between quasi-maximum likelihood estimation (QMLE) and maximum likelihood estimate (MLE) consider the GARCH(1,1)-$\tilde{t}(d)$ model with leverage. We have

$$R_{PF,t+1} = \sigma_{PF,t+1} z_{t+1}, \text{ with } z_{t+1} \overset{i.i.d.}{\sim} \tilde{t}(d), \text{ where}$$
$$\sigma^2_{PF,t+1} = \omega + \alpha \left(R_{PF,t} - \theta\sigma_{PF,t}\right)^2 + \beta\sigma^2_{PF,t}$$

We can estimate all the parameters $\{\omega, \alpha, \beta, \theta, d\}$ in one step using $\ln L_2$ from before, which would correspond to exact MLE. Alternatively, we can first estimate the GARCH parameters $\{\omega, \alpha, \beta, \theta\}$ using the QMLE method in Chapter 4, which assumes the likelihood from a normal distribution, and then estimate the conditional distribution parameter, d, from $\ln L_1$. In this simple example, exact MLE is clearly feasible as the total number of parameters is only five.

6.2 An Easy Estimate of d

While the maximum likelihood estimation outlined here has nice properties, there is a very simple alternative estimation procedure available for the t distribution. If the conditional variance model has already been estimated, then we are only estimating one parameter, namely d. Because there is a simple closed-form relationship between

d and the excess kurtosis, ζ_2, this suggests first simply calculating ζ_2 from the z_t variable and then calculating d from

$$\zeta_2 = 6/(d-4) \Rightarrow d = 6/\zeta_2 + 4$$

Thus, if excess kurtosis is found to be 1, for example, then the estimate of d is 10. This is an example of a method-of-moments estimate, where we match the fourth sample moment of the data (in this case z_t) to the fourth moment from the assumed distribution (in this case the t distribution). Notice that this estimate of d is conditional on having estimated the GARCH parameters in a previous step using QMLE. Only when the GARCH parameters have been estimated on returns can we define the time series of GARCH shocks, z_t.

6.3 Calculating Value-at-Risk and Expected Shortfall

Once d is estimated, we can calculate the VaR for the portfolio return

$$R_{PF,t+1} = \sigma_{PF,t+1} z_{t+1}, \quad \text{with } z_{t+1} \overset{i.i.d.}{\sim} \tilde{t}(d)$$

as

$$VaR^p_{t+1} = -\sigma_{PF,t+1} \tilde{t}_p^{-1}(d)$$

where $\tilde{t}_p^{-1}(d)$ is the pth quantile of the $\tilde{t}(d)$ distribution.

Thus, we have

$$VaR^p_{t+1} = -\sigma_{PF,t+1} \sqrt{\frac{d-2}{d}} t_p^{-1}(d)$$

where we have used the below result relating the quantiles of the standardized $\tilde{t}(d)$ distribution to that of the conventional Student's $t(d)$.

The formula for the expected shortfall is

$$ES^p_{t+1} = -\sigma_{PF,t+1} ES_{\tilde{t}(d)}(p), \text{ where}$$
$$ES_{\tilde{t}(d)}(p) = \frac{C(d)}{p} \left[\left[1 + \frac{1}{d-2} t_p^{-1}(d) \right]^{\frac{1-d}{2}} \frac{d-2}{1-d} \right], \text{ with}$$
$$C(d) = \frac{\Gamma((d+1)/2)}{\Gamma(d/2)\sqrt{\pi(d-2)}} \text{ as before.}$$

Appendix A at the end of this chapter gives the derivation of $ES_{\tilde{t}(d)}(p)$.

6.4 QQ Plots

We can generalize the preceding QQ plot to assess the appropriateness of nonnormal distributions as well. In particular, we would like to assess if the returns standardized by the GARCH model conform to the $\tilde{t}(d)$ distribution.

However, the quantile of the standardized $\tilde{t}(d)$ distribution is usually not easily found in software packages, whereas the quantile from the conventional Student's $t(d)$ distribution is. We therefore need the relationship

$$\Pr\left(z_t\sqrt{\frac{d}{d-2}} < t_p^{-1}(d)\right) = p$$

$$\Leftrightarrow \Pr\left(z_t < t_p^{-1}(d)\sqrt{\frac{d-2}{d}}\right) = p$$

$$\Leftrightarrow \tilde{t}_p^{-1}(d) = \sqrt{\frac{d-2}{d}}t_p^{-1}(d)$$

where $t_p^{-1}(d)$ is the pth quantile of the conventional Student's $t(d)$ distribution.

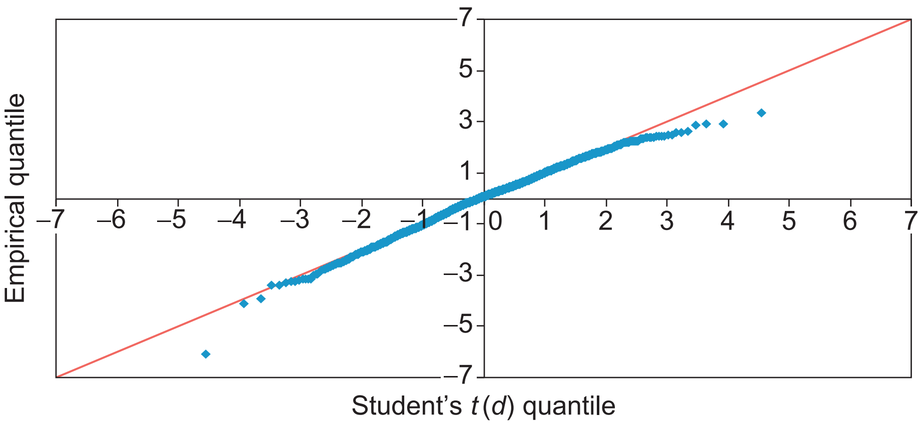
We are now ready to construct the QQ plot as

$$\{X_i, Y_i\} = \left\{\sqrt{\frac{d-2}{d}}t_{(i-0.5)/T}^{-1}(d), z_i\right\}$$

where z_i again denotes the ith sorted standardized return.

Figure 6.3 shows the QQ plot of the standardized returns from the GARCH-$\tilde{t}(d)$ with leverage, estimated using QMLE. d is estimated to be 11.4. Notice that the t distribution fits the left tail better than the normal distribution, but this happens partly at the cost of fitting the right tail worse.

Figure 6.3 QQ plot of S&P 500 GARCH shocks against the standardized t distribution.



Notes: We scatter plot the empirical quantiles of the S&P 500 GARCH shocks against the quantiles of the standardized Student's t distribution. The red line has a slope of one.

The symmetry of the $\tilde{t}(d)$ distribution appears to be somewhat at odds with this particular data set. We therefore next consider a generalization of the t distribution that allows for asymmetry.

7 The Asymmetric *t* Distribution

The Student's t distribution can allow for kurtosis in the conditional distribution but not for skewness. It is possible, however, to develop a generalized, asymmetric version of the Student's t distribution. It is defined by pasting together two distributions at a point $-A/B$ on the horizontal axis. The density function is defined by

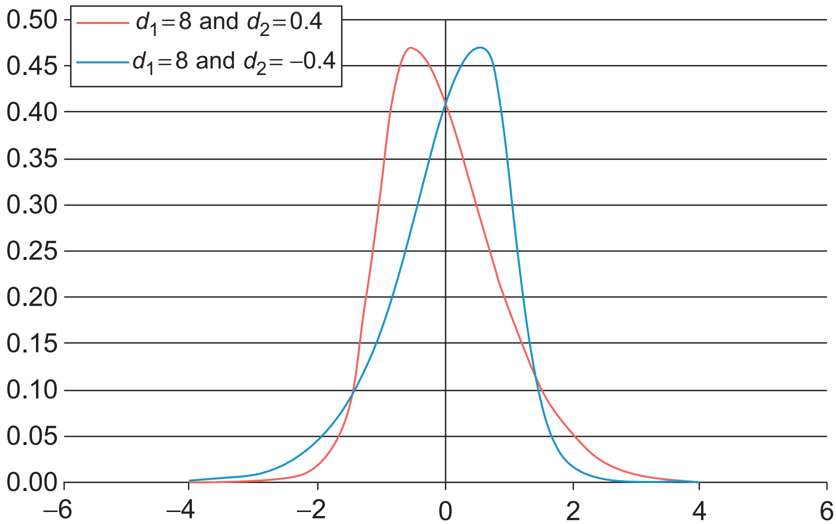
$$f_{asyt}(z; d_1, d_2) = \begin{cases} BC\left[1+(Bz+A)^2 / \left((1-d_2)^2 (d_1-2)\right)\right]^{-(1+d_1)/2}, & \text{if } z < -A/B \\ BC\left[1+(Bz+A)^2 / \left((1+d_2)^2 (d_1-2)\right)\right]^{-(1+d_1)/2}, & \text{if } z \geq -A/B \end{cases}$$

where

$$A = 4d_2C\frac{d_1-2}{d_1-1}, \quad B = \sqrt{1+3d_2^2-A^2}, \quad C = \frac{\Gamma((d_1+1)/2)}{\Gamma(d_1/2)\sqrt{\pi(d_1-2)}}$$

and where $d_1 > 2$, and $-1 < d_2 < 1$. Note that $C(d_1) = C(d)$ from the symmetric Student's t distribution. Figure 6.4 shows the asymmetric t distribution for $\{d_1, d_2\} = \{8, -0.4\}$ in blue, and $\{d_1, d_2\} = \{8, +0.4\}$ in red.

Figure 6.4 The asymmetric t distribution.



Notes: The red line plots the asymmetric t distribution with $d_2 = +0.4$, which implies a skewness of +1. The blue line corresponds to $d_2 = -0.4$, which implies a skewness of −1. The d_2 parameter is set to 8 in both cases, which implies an excess kurtosis of 2.6.

In order to derive the moments of the distribution we first define

$$m_2 = 1 + 3d_2^2$$

$$m_3 = 16Cd_2\left(1 + d_2^2\right)\frac{(d_1 - 2)^2}{(d_1 - 1)(d_1 - 3)}, \quad \text{for } d_1 > 3$$

$$m_4 = 3\frac{(d_1 - 2)}{(d_1 - 4)}\left(1 + 10d_2^2 + 5d_2^4\right), \quad \text{for } d_1 > 4$$

With these in hand, we can derive the first four moments of the asymmetric t distribution to be

$$\mu \equiv E[z] = 0$$

$$\sigma^2 \equiv E[(z - E[z])^2] = 1$$

$$\zeta_1 \equiv E[z^3]/\sigma^3 = \left[m_3 - 3Am_2 + 2A^3\right]/B^3$$

$$\zeta_2 \equiv E[z^4]/\sigma^4 - 3 = \left[m_4 - 4Am_3 + 6A^2m_2 - 3A^4\right]/B^4 - 3$$

Note from the formulas that although skewness is zero if d_2 is zero, skewness and kurtosis are generally highly nonlinear functions of d_1 and d_2.

Consider again the two distributions in Figure 6.4. The red line corresponds to a skewness of $+1$ and an excess kurtosis of 2.6; the blue line corresponds to a skewness of -1 and an excess kurtosis of 2.6.

Skewness and kurtosis are both functions of d_1 as well as d_2. The upper panel of Figure 6.5 shows skewness plotted as a function of d_2 on the horizontal axis. The blue line uses $d_1 = 5$ (high kurtosis) and the red line uses $d_1 = 10$ (moderate kurtosis). The lower panel of Figure 6.5 shows kurtosis plotted as a function of d_1 on the horizontal axis. The red line uses $d_2 = 0$ (no skewness) and the blue line uses $d_2 = 0.5$ (positive skewness). The asymmetric t distribution is capable of generating a wide range of skewness and kurtosis levels.

Notice that the symmetric standardized Student's t is a special case of the asymmetric t where $d_1 = d$, $d_2 = 0$, which implies $A = 0$ and $B = 1$, so we get

$$m_2 = 1, m_3 = 0, m_4 = 3\frac{(d - 2)}{(d - 4)}$$

which yields

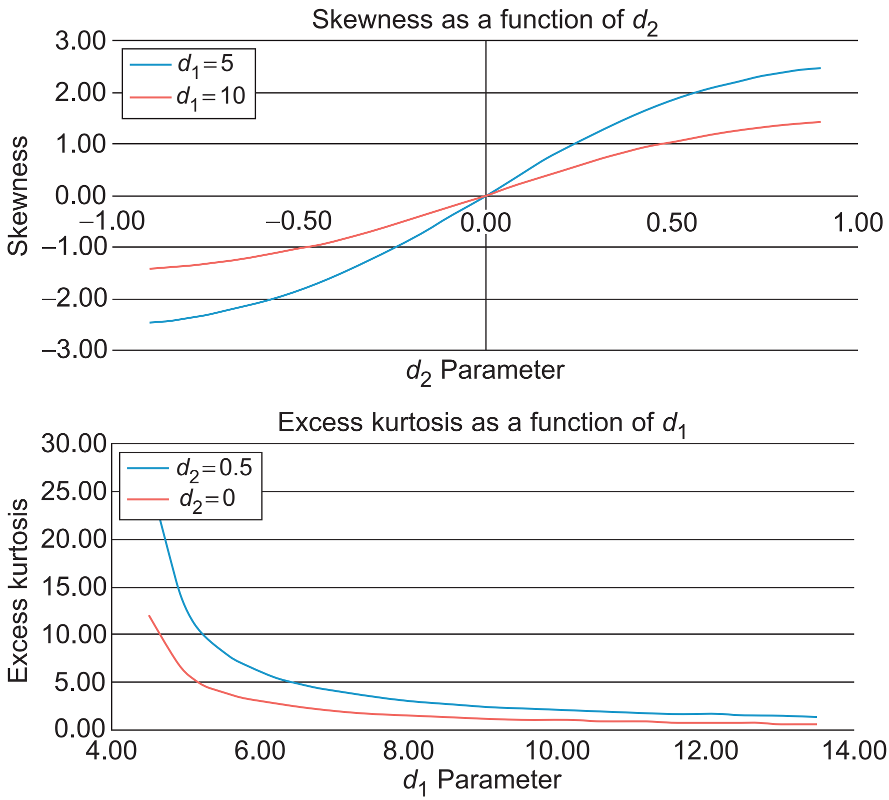
$$\zeta_1 = 0, \text{and } \zeta_2 = 3\frac{(d - 2)}{(d - 4)} - 3 = 6/(d - 4)$$

as in the previous section.

7.1 Estimation of d_1 and d_2

The parameters d_1 and d_2 in the asymmetric t distribution can be estimated via maximum likelihood as before. The only added complication is that the shape of the

Figure 6.5 Skewness and kurtosis in the asymmetric *t* distribution.



Notes: In the top panel we plot skewness in the asymmetric *t* distribution as a function of the d_2 parameter. Skewness is also a function of d_1. The blue line uses $d_1 = 5$ and the red line uses $d_1 = 10$. In the bottom panel we plot excess kurtosis as a function of d_1. Excess kurtosis is also a function of d_2. The red line uses $d_2 = 0$ and the blue line uses $d_2 = 0.5$.

likelihood function on any given day will depend on the value of the shock z_t. As before we can define the likelihood function for z_t as

$$\ln L_1 = \sum_{t=1}^{T} \ln(f_{asyt}(z_t; d_1, d_2))$$

where

$$\ln(f_{asyt}(z_t; d_1, d_2)) = \begin{cases} \ln(BC) - \frac{(1+d_1)}{2} \ln\left(\left[1 + \frac{(Bz_t+A)^2}{((1-d_2)^2(d_1-2))}\right]\right), & \text{if } z_t < -A/B \\ \ln(BC) - \frac{(1+d_1)}{2} \ln\left(\left[1 + \frac{(Bz_t+A)^2}{((1+d_2)^2(d_1-2))}\right]\right), & \text{if } z_t \geq -A/B \end{cases}$$

This estimation assumes that the conditional variance is estimated without error so that we can treat $z_t = R_{PF,t}/\sigma_{PF,t}$ as a regular data point. Alternatively joint estimation

of the volatility and distribution parameters can be done using

$$\ln L_2 = \sum_{t=1}^{T} \ln(f_{asyt}(R_{PF,t}; d_1, d_2)) = \ln L_1 - \sum_{t=1}^{T} \ln(\sigma^2_{PF,t})/2$$

as before.

We can also estimate d_1 and d_2 using sample estimates of skewness, ζ_1, and kurtosis, ζ_2. Unfortunately, the relationship between the parameters and the moments is nonlinear and so the equations

$$\zeta_1 = \left[m_3 - 3Am_2 + 2A^3\right]/B^3$$
$$\zeta_2 = \left[m_4 - 4Am_3 + 6A^2m_2 - 3A^4\right]/B^4 - 3$$

must be solved numerically to get moment-based estimates of d_1 and d_2 using the formulas for A, B, m_2, m_3, and m_4, earlier.

7.2 Calculating Value-at-Risk and Expected Shortfall

Once d_1 and d_2 are estimated, we can calculate the Value-at-Risk for the portfolio return

$$R_{PF,t+1} = \sigma_{PF,t+1} z_{t+1}, \quad \text{with } z_{t+1} \overset{i.i.d.}{\sim} F_{asyt}(d_1, d_2)$$

as

$$VaR^p_{t+1} = -\sigma_{PF,t+1} F^{-1}_{asyt}(p; d_1, d_2)$$

where $F^{-1}_{asyt}(p; d_1, d_2)$ is the pth quantile of the asymmetric t distribution, which is given by

$$F^{-1}_{asyt}(p; d_1, d_2) = \begin{cases} \frac{1}{B}\left[(1-d_2)\sqrt{\frac{d_1-2}{d_1}}\, t^{-1}_{p/(1-d_2)}(d_1) - A\right], & \text{if } p < \frac{1-d_2}{2} \\ \frac{1}{B}\left[(1+d_2)\sqrt{\frac{d_1-2}{d_1}}\, t^{-1}_{(p+d_2)/(1+d_2)}(d_1) - A\right], & \text{if } p \geq \frac{1-d_2}{2} \end{cases}$$

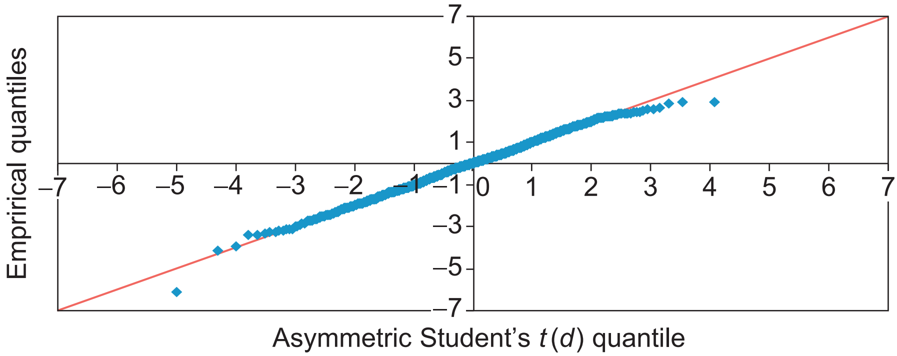
where we have used the inverse of the symmetric t distribution, $t^{-1}_p(d)$, for different values of p and d.

The expected shortfall can be computed as

$$ES^p_{t+1} = -\sigma_{PF,t+1} ES_{asyt}(p)$$

where the formula for $ES_{asyt}(p)$ is a complicated function of d_1 and d_2 and is given in Appendix A at the end of this chapter.

Figure 6.6 QQ plot of S&P 500 GARCH shocks against the asymmetric *t* distribution.



Notes: We scatter plot the empirical quantiles of the S&P 500 GARCH shocks against the quantiles of the asymmetric Student's *t* distribution. The red line has a slope of one.

7.3 QQ Plots

Armed with the earlier formula for the inverse cumulative density function (CDF) we can again construct the QQ plot as

$$\{X_i, Y_i\} = \left\{F_{asyt}^{-1}((i-0.5)/T; d_1, d_2), z_i\right\}$$

where z_i again denotes the ith sorted standardized return.

Figure 6.6 shows the QQ plot for the asymmetric *t* distribution. Note that the asymmetric *t* distribution is able to fit the S&P 500 shocks quite well. Only the single largest negative shock seems to deviate substantially from the 45-degree line.

In conclusion, the asymmetric *t* distribution is somewhat cumbersome to estimate and implement but it is capable of fitting GARCH shocks from daily asset returns quite well.

The *t* distributions—and any other distribution—attempt to fit the entire range of outcomes using all the data available. Consequently, the estimated parameters in the distribution (for example d_1 and d_2) may be influenced excessively by data values close to zero, of which we observe many but of which risk managers care little about. We therefore now turn to an alternative approach that only makes use of the extreme return observations that of course contain crucial information for risk management.

8 Extreme Value Theory (EVT)

Typically, the biggest risks to a portfolio is the sudden occurrence of a single large negative return. Having explicit knowledge of the probabilities of such extremes is, therefore, at the essence of financial risk management. Consequently, risk managers ought to focus on modeling the tails of the returns distribution. Fortunately, a branch of statistics is devoted exactly to the modeling of such extreme values.

The central result in extreme value theory states that the extreme tail of a wide range of distributions can approximately be described by a relatively simple distribution, the so-called Generalized Pareto Distribution (GPD).

Virtually all results in extreme value theory (EVT) assume that returns are i.i.d. and therefore are not very useful unless modified to the asset return environment. Asset returns appear to approach normality at long horizons, thus EVT is more important at short horizons, such as daily. Unfortunately, the i.i.d. assumption is the least appropriate at short horizons due to the time-varying variance patterns. Therefore we need to get rid of the variance dynamics before applying EVT. Consider again, therefore, the standardized portfolio returns

$$z_{t+1} = R_{PF,t+1}/\sigma_{PF,t+1} \overset{i.i.d.}{\sim} D(0,1)$$

Fortunately, it is typically reasonable to assume that these standardized returns are i.i.d. Thus, we will proceed to apply EVT to the standardized returns and then combine EVT with the variance models estimated in Chapters 4 and 5 in order to calculate *VaR*s.

8.1 The Distribution of Extremes

Consider the entire distribution of the shocks, z_t, as illustrated for example by the histogram in Figure 6.1. EVT is concerned only with the tail of the distribution and we first have to decide what we mean by the tail. To this end define a threshold value u on the horizontal axis of the histogram. The threshold could for example be set to 0.02 in the top panel of Figure 6.1.

The key result in extreme value theory states that as you let the threshold u go to infinity, in almost any distribution you can think of, the distribution of observations beyond the threshold (call them y) converge to the Generalized Pareto Distribution, $GPD(y;\xi,\beta)$, where

$$GPD(y;\xi,\beta) = \begin{cases} 1-(1+\xi y/\beta)^{-1/\xi} & \text{if } \xi > 0 \\ 1-\exp(-y/\beta) & \text{if } \xi = 0 \end{cases}$$

with $\beta > 0$ and $y \geq u$. The so-called tail-index parameter ξ is key as it controls the shape of the distribution tail and in particular how quickly the tail goes to zero when the extreme, y, goes to infinity.

Standard distributions that are covered by the EVT result include those that are heavy tailed, for example the Student's $t(d)$ distribution, where the tail-index parameter, ξ, is positive. This is, of course, the case of most interest in financial risk management, where returns tend to have fat tails.

The normal distribution is also covered. We noted earlier that a key difference between the Student's $t(d)$ distribution and the normal distribution is that the former has power tails and the latter has exponential tails. Thus, for the normal distribution we have that the tail parameter, ξ, equals zero.

Finally, thin-tailed distributions are covered when the tail parameter $\xi < 0$, but they are not relevant for risk management and so we will not consider that case here.

8.2 Estimating the Tail Index Parameter, ξ

We could use MLE to estimate the GPD distribution defined earlier. However, if we are willing to assume that the tail parameter, ξ, is strictly positive, as is typically the case in risk management, then a very easy estimator exists, namely the so-called Hill estimator. The idea behind the Hill estimator is to approximate the GPD distribution by

$$F(y) = 1 - cy^{-1/\xi} \approx 1 - (1 + \xi y/\beta)^{-1/\xi} = GPD\,(y; \xi, \beta)$$

for $y > u$ and $\xi > 0$. Recall now the definition of a conditional distribution,

$$f\,(y|y > u) = f\,(y)\,/\,\Pr\,(y > u) = f\,(y)\,/\,(1 - F(u))\,, \quad \text{for } y > u$$

Note that from the definition of $F(y)$ we have

$$F\,(u) = 1 - cu^{-1/\xi}$$

We can also get the density function of y from $F(y)$:

$$f(y) = \frac{\partial F(y)}{\partial y} = \frac{1}{\xi} c y^{-1/\xi - 1}$$

We are now ready to construct the likelihood function for all observations y_i larger than the threshold, u, as

$$L = \prod_{i=1}^{T_u} f(y_i)/(1 - F(u)) = \prod_{i=1}^{T_u} \frac{1}{\xi} c y_i^{-1/\xi - 1}/(cu^{-1/\xi}), \quad \text{for } y_i > u$$

where T_u is the number of observations y larger than u. The log-likelihood function is therefore

$$\ln L = \sum_{i=1}^{T_u} \left(-\ln(\xi) - (1/\xi + 1)\ln(y_i) + \frac{1}{\xi}\ln(u) \right)$$

Taking the derivative with respect to ξ and setting it to zero yields the Hill estimator of the tail index parameter

$$\xi = \frac{1}{T_u} \sum_{i=1}^{T_u} \ln(y_i/u)$$

We can estimate the c parameter by ensuring that the fraction of observations beyond the threshold is accurately captured by the density as in

$$F(u) = 1 - cu^{-1/\xi} = 1 - T_u/T$$

Solving this equation for c yields the estimate

$$c = \frac{T_u}{T} u^{1/\xi}$$

Our estimate of the cumulative density function for observations beyond u is, therefore

$$F(y) = 1 - cy^{-1/\xi} = 1 - \frac{T_u}{T}(y/u)^{-1/\xi}$$

Notice that our estimates are available in closed form—they do not require numerical optimization. They are, therefore, extremely easy to calculate.

So far we have implicitly referred to extreme returns as being large gains. Of course, as risk managers we are more interested in extreme negative returns corresponding to large losses. To this end, we simply do the EVT analysis on the negative of returns instead of returns themselves.

8.3 Choosing the Threshold, u

Until now, we have focused on the benefits of the EVT methodology, such as the explicit focus on the tails, and the ability to study each tail separately, thereby avoiding unwarranted symmetry assumptions. The EVT methodology does have an Achilles heel however, namely the choice of threshold, u. When choosing u we must balance two evils: bias and variance. If u is set too large, then only very few observations are left in the tail and the estimate of the tail parameter, ξ, will be very noisy. If on the other hand u is set too small, then the EVT theory may not hold, meaning that the data to the right of the threshold does not conform sufficiently well to the Generalized Pareto Distribution to generate unbiased estimates of ξ.

Simulation studies have shown that in typical data sets with daily asset returns, a good rule of thumb is to set the threshold so as to keep the largest 50 observations for estimating ξ; that is, we set $T_u = 50$. Visually gauging the QQ plot can provide useful guidance as well. Only those observations in the tail that are clearly deviating from the 45-degree line indicating the normal distribution should be used in the estimation of the tail index parameter, ξ.

8.4 Constructing the QQ Plot from EVT

We next want to show the QQ plot of the large losses using the EVT distribution. Define y to be a standardized loss; that is,

$$y_i = -R_{PF,i}/\sigma_{PF,i}$$

The first step is to estimate ξ and c from the losses, y_i, using the Hill estimator from before.

Next, we need to compute the inverse cumulative distribution function, which gives us the quantiles. Recall the EVT cumulative density function from before:

$$F(y) = 1 - cy^{-1/\xi} = 1 - \frac{T_u}{T}(y/u)^{-1/\xi}$$

We now set the estimated cumulative probability function equal to $1-p$ so that there is only a p probability of getting a standardized loss worse than the quantile, F_{1-p}^{-1}, which is implicitly defined by

$$F(F_{1-p}^{-1}) = 1 - p$$

From the definition of $F(\bullet)$, we can solve for the quantile to get

$$F_{1-p}^{-1} = u[p/(T_u/T)]^{-\xi}$$

We are now ready to construct the QQ plot from EVT using the relationship

$$\{X_i, Y_i\} = \left\{u[\{(i-0.5)/T\}/(T_u/T)]^{-\xi}, y_i\right\}$$

where y_i is the ith sorted standardized loss.

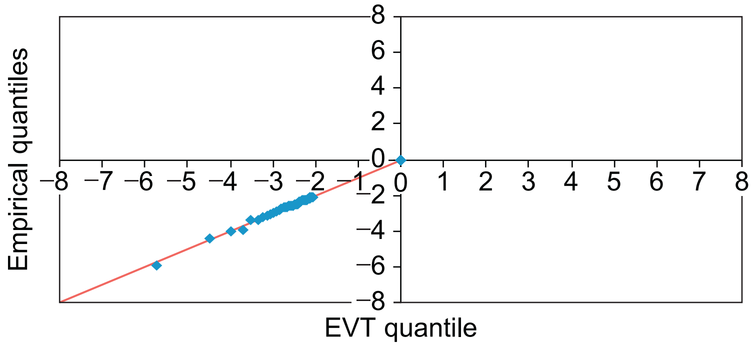
Figure 6.7 shows the QQ plots of the EVT tails for large losses from the standardized S&P 500 returns. For this data, ξ is estimated to be 0.22.

8.5 Calculating VaR and ES from the EVT Quantile

We are, of course, ultimately interested not in QQ plots but rather in portfolio risk measures such as Value-at-Risk. Using again the loss quantile F_{1-p}^{-1} defined earlier by

$$F_{1-p}^{-1} = u[p/(T_u/T)]^{-\xi}$$

Figure 6.7 QQ plot of daily S&P 500 tail shocks against the EVT distribution.



Notes: We plot the quantiles of the largest negative S&P 500 GARCH shocks against the quantiles of the EVT distribution. The line has a slope of one.

the *VaR* from the EVT combined with the variance model is now easily calculated as

$$VaR_{t+1}^{p} = \sigma_{PF,t+1} F_{1-p}^{-1} = \sigma_{PF,t+1} u \left[p/\left(T_u/T\right)\right]^{-\xi}$$

The reason for using the $(1-p)$th quantile from the EVT loss distribution in the *VaR* with coverage rate p is that the quantile such that $(1-p)\cdot 100\%$ of *losses* are smaller than it is the same as minus the quantile such that $p\cdot 100\%$ of *returns* are smaller than it.

We usually calculate the *VaR* taking Φ_p^{-1} to be the pth quantile from the standardized *return* so that

$$VaR_{t+1}^{p} = -\sigma_{PF,t+1} \Phi_p^{-1}$$

But we now take F_{1-p}^{-1} to be the $(1-p)$th quantile of the standardized *loss* so that

$$VaR_{t+1}^{p} = \sigma_{PF,t+1} F_{1-p}^{-1}$$

The expected shortfall can be computed using

$$ES_{t+1}^{p} = \sigma_{PF,t+1} ES_{EVT}\left(p\right)$$

where

$$ES_{EVT}\left(p\right) = -\frac{u}{\xi - 1}\left[p/\left(T_u/T\right)\right]^{-\xi}$$

when $\xi < 1$. This expression is derived in Appendix C.

In general, the ratio of *ES* to *VaR* for fat-tailed distribution will be higher than that of the normal. When using the Hill approximation of the EVT tail the previous formulas for *VaR* and *ES* show that we have a particularly simple relationship, namely

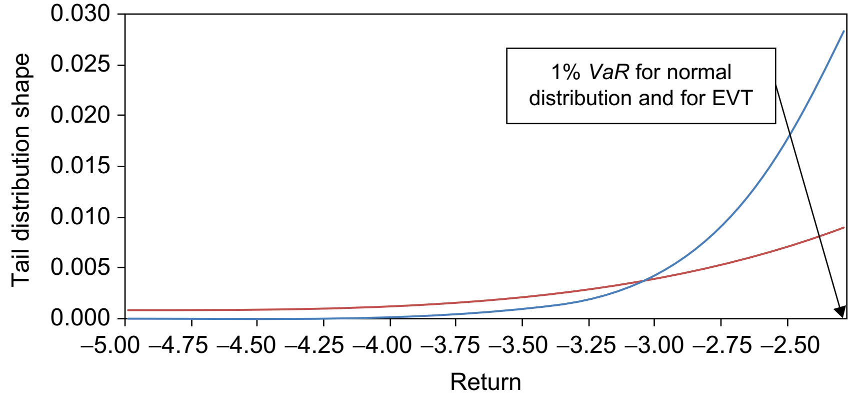
$$\frac{ES_{t+1}^{p}}{VaR_{t+1}^{p}} = \frac{1}{1-\xi}$$

so that for fat-tailed distributions where $\xi > 0$, the fatter the tail, the larger the ratio of *ES* to *VaR*.

In Figure 6.8 we plot the tail shape of a normal distribution (the blue line) and EVT distribution (red line) where $\xi = 0.5$. The plot has been constructed so that the 1% *VaR* is 2.33 in both distributions. The probability mass under the two curves is therefore 1% in both cases. Note however, that the risk profile is very different. The normal distribution has a tail that goes to a virtual zero very quickly as the losses get extreme. The EVT distribution on the other hand implies a nontrivial probability of getting losses in excess of five standard deviations.

The preceding formula shows that when $\xi = 0.5$ then the *ES* to *VaR* ratio is 2. Thus even though the 1% *VaR* is the same in the two distributions by construction, the *ES* measure reveals the differences in the risk profiles of the two distributions, which arises from one being fat-tailed. The *VaR* does not reveal this difference unless the *VaR* is reported for several extreme coverage probabilities, p.

Figure 6.8 Tail shapes of the normal distribution (blue) and EVT (red).



Notes: We plot the tail shape of the standard normal distribution in blue and the tail shape of an EVT distribution with tail index parameter of 0.5 in red. Both distributions have a 1% *VaR* of 2.33.

9 Summary

Time-varying variance models help explain nonnormal features of financial returns data. However, the distribution of returns standardized by a dynamic variance tends to be fat-tailed and may be skewed. This chapter has considered methods for modeling the nonnormality of portfolio returns by building on the variance and correlation models established in earlier chapters and using the same maximum likelihood estimation techniques.

We have introduced a graphical tool for visualizing nonnormality in the data, the so-called QQ plot. This tool was used to assess the appropriateness of alternative distributions.

Several alternative approaches were considered for capturing nonnormality in the portfolio risk distribution.

- The Filtered Historical Simulation approach, which uses the empirical distribution of the GARCH shocks and avoids making specific distribution choices
- The Cornish-Fisher approximation to the shock distribution, which allows for skewness and kurtosis using the sample moments that are easily estimated
- The standardized t distribution, which allows for fatter tails than the normal, but assumes that the distribution is symmetric around zero
- The asymmetric t distribution, which is more complex but allows for skewness as well as kurtosis
- Extreme value theory, which models the tail of the distribution directly using only extreme shocks in the sample

This chapter has focused on one-day-ahead distribution modeling. The multiday distribution requires Monte Carlo simulation, which will be covered in Chapter 8.

We end this chapter by stressing that in Part II of the book we have analyzed the conditional distribution of the aggregate portfolio return only. Thus, the distribution is dependent on the particular set of current portfolio weights, and the distribution must be reestimated when the weights change. Part III of the book presents multivariate risk models where portfolio weights can be rebalanced without requiring reestimation of the model.

Appendix A: *ES* for the Symmetric and Asymmetric *t* Distributions

In this appendix we derive the expected shortfall (*ES*) measure for the asymmetric *t* distribution. The *ES* for the symmetric case will be given as a special case at the end.

We want to compute $ES_{asyt}(p)$. Let us assume for simplicity that p is such that $Q \equiv F_{asyt}^{-1}(p; d_1, d_2) < -\frac{A}{B}$, then

$$ES_{asyt}(p) = \frac{BC}{p} \int_{-\infty}^{Q} z \left[1 + \frac{1}{d_1 - 2} \left(\frac{Bz + A}{1 - d_2} \right)^2 \right]^{-\frac{d_1+1}{2}} dz$$

We use the change of variable

$$x = \frac{Bz + A}{1 - d_2}$$

$$dx = \frac{B}{1 - d_2} dz$$

which yields

$$\begin{aligned} ES_{asyt}(p) &= \frac{C(1-d_2)}{Bp} \int_{-\infty}^{\frac{BQ+A}{1-d_2}} (x(1-d_2) - A) \left[1 + \frac{1}{d_1 - 2} x^2 \right]^{-\frac{d_1+1}{2}} dx \\ &= \frac{C(1-d_2)^2}{Bp} \int_{-\infty}^{\frac{BQ+A}{1-d_2}} x \left[1 + \frac{1}{d_1 - 2} x^2 \right]^{-\frac{d_1+1}{2}} dx \\ &\quad - \frac{AC(1-d_2)}{Bp} \int_{-\infty}^{\frac{BQ+A}{1-d_2}} \left[1 + \frac{1}{d_1 - 2} x^2 \right]^{-\frac{d_1+1}{2}} dx \end{aligned}$$

The first integral can be solved to get

$$\frac{C(1-d_2)^2}{Bp}\left[\left[1+\frac{1}{d_1-2}x^2\right]^{\frac{1-d_1}{2}}\frac{d_1-2}{1-d_1}\right]_{-\infty}^{\frac{BQ+A}{1-d_2}}$$
$$=\frac{C(1-d_2)^2}{Bp}\left[\left[1+\frac{1}{d_1-2}\left(\frac{BQ+A}{1-d_2}\right)\right]^2\right]^{\frac{1-d_1}{2}}\frac{d_1-2}{1-d_1}$$

and the second integral can be related to the regular symmetric Student's t distribution by

$$-\frac{AC(1-d_2)}{Bp}\frac{\sqrt{\pi(d_1-2)}\Gamma\left(\frac{d_1}{2}\right)}{\Gamma\left(\frac{d_1+1}{2}\right)}t_{d_1}\left(\sqrt{\frac{d_1}{d_1-2}}\frac{BQ+A}{1-d_2}\right)$$

where t_{d_1} is the CDF of a Student's t distribution with d_1 degrees of freedom.

Therefore,

$$ES_{asyt}(p)=\frac{C(1-d_2)^2}{Bp}\left[\left[1+\frac{1}{d_1-2}\left(\frac{BQ+A}{1-d_2}\right)^2\right]^{\frac{1-d_1}{2}}\frac{d_1-2}{1-d_1}\right]$$
$$-\frac{AC(1-d_2)}{Bp}\frac{\sqrt{\pi(d_1-2)}\Gamma\left(\frac{d_1}{2}\right)}{\Gamma\left(\frac{d_1+1}{2}\right)}t_{d_1}\left(\sqrt{\frac{d_1}{d_1-2}}\frac{BQ+A}{1-d_2}\right)$$

In the symmetric case we have $d_1=d$, $d_2=0$, $A=0$, and $B=1$ and so we get

$$ES_{\tilde{t}(d)}(p)=\frac{C}{p}\left[\left[1+\frac{1}{d-2}Q^2\right]^{\frac{1-d}{2}}\frac{d-2}{1-d}\right]$$

where now $Q\equiv t_p^{-1}(d)$.

Appendix B: Cornish-Fisher *ES*

The Cornish-Fisher approach assumes an approximate distribution of the form

$$f(z)=\phi(z)\left\{1+\frac{\zeta_1}{6}\left(z^3-3z\right)+\frac{\zeta_2}{24}\left(z^4-6z^2+3\right)\right\}$$

The expected shortfall is again defined as

$$ES_{CF}(p) = \frac{1}{p}\int_{-\infty}^{Q} z\phi(z)\left\{1 + \frac{\zeta_1}{6}\left(z^3 - 3z\right) + \frac{\zeta_2}{24}\left(z^4 - 6z^2 + 3\right)\right\}dz$$

where $Q = CF_p^{-1}$. Solving the integral we get

$$\begin{aligned}
ES_{CF}(p) &= \frac{1}{p}\int_{-\infty}^{Q} z\phi(z)\left\{1 + \frac{\zeta_1}{6}\left(z^3 - 3z\right) + \frac{\zeta_2}{24}\left(z^4 - 6z^2 + 3\right)\right\}dz \\
&= \frac{1}{p}\left\{\int_{-\infty}^{Q} z\phi(z)\,dz + \frac{\zeta_1}{6}\int_{-\infty}^{Q}\left(z^4 - 3z^2\right)\phi(z)\,dz \right. \\
&\qquad \left. + \frac{\zeta_2}{24}\int_{-\infty}^{Q}\left(z^5 - 6z^3 + 3z\right)\phi(z)\,dz\right\} \\
&= \frac{1}{p}\left\{[-\phi(z)]_{-\infty}^{Q} + \frac{\zeta_1}{6}\left[-z^3\phi(z)\right]_{-\infty}^{Q} + \frac{\zeta_2}{24}\left[(-z^4 + 2z^2 + 1)\phi(z)\right]_{-\infty}^{Q}\right\} \\
&= \frac{1}{p}\left\{-\phi(Q) - \frac{\zeta_1}{6}Q^3\phi(Q) + \frac{\zeta_2}{24}(-Q^4 + 2Q^2 + 1)\phi(Q)\right\} \\
&= \frac{-\phi(Q)}{p}\left\{1 + \frac{\zeta_1}{6}Q^3 + \frac{\zeta_2}{24}(Q^4 - 2Q^2 - 1)\right\}
\end{aligned}$$

Appendix C: Extreme Value Theory *ES*

Expected shortfall in the Hill approximation to EVT can be derived as

$$\begin{aligned}
ES_{EVT}(p) &= \frac{1}{p}\int_{F_{1-p}^{-1}}^{\infty} y\,d\left[1 - \frac{T_u}{T}(y/u)^{-1/\xi}\right] \\
&= \frac{1}{p}\int_{F_{1-p}^{-1}}^{\infty} y\frac{T_u u^{1/\xi}}{\xi T}y^{-\frac{1}{\xi}-1}dy = \frac{T_u u^{1/\xi}}{p\xi T}\int_{F_{1-p}^{-1}}^{\infty} y^{-\frac{1}{\xi}}dy \\
&= \frac{T_u u^{1/\xi}}{p\xi T}\left[\frac{y^{-\frac{1}{\xi}+1}}{-\frac{1}{\xi}+1}\right]_{F_{1-p}^{-1}}^{\infty} = \frac{T_u u^{1/\xi}}{p(\xi - 1)T}\left[-(F_{1-p}^{-1})^{-\frac{1}{\xi}+1}\right] \\
&= \frac{T_u u^{1/\xi}}{p(\xi - 1)T}\left[-\left(u[p/(T_u/T)]^{-\xi}\right)^{-\frac{1}{\xi}+1}\right]
\end{aligned}$$

$$= -\frac{u}{\xi - 1}\left(\frac{T_u}{pT}\right)^{\xi} = -\frac{u}{\xi - 1}\left(p/\left(T_u/T\right)\right)^{-\xi}$$

Further Resources

Details on the asymmetric t distribution considered here can be found in Hansen (1994), Fernandez and Steel (1998), and Jondeau and Rockinger (2003). Hansen (1994) and Jondeau and Rockinger (2003) also discuss time-varying skewness and kurtosis models. The GARCH-$\tilde{t}(d)$ model was introduced by Bollerslev (1987).

Applications of extreme value theory to financial risk management is discussed in McNeil (1999). The choice of threshold value in the GARCH-EVT model is discussed in McNeil and Frey (2000). Huisman, Koedijk, Kool, and Palm (2001) explore improvements to the simple Hill estimator considered here. McNeil (1997) and McNeil and Saladin (1997) discuss the use of QQ plots in deciding on the threshold parameter, u. Brooks, Clare, Molle, and Persand (2005) compare various EVT approaches.

Multivariate extensions to the univariate EVT analysis considered here can be found in Longin (2000), Longin and Solnik (2001), and Poon, Rockinger, and Tawn (2003).

The expected shortfall measure for the Cornish-Fisher approximation is developed in Giamouridis (2006). In the spirit of the Cornish-Fisher approach, Jondeau and Rockinger (2001) develop a Gram-Charlier approach to return distribution modeling.

Many alternative conditional distribution approaches exist. Kuerster et al. (2006) perform a large-scale empirical study.

GARCH and RV models can also be combined with jump processes. See Maheu and McCurdy (2004), Ornthanalai (2010), and Christoffersen, Jacobs, and Ornthanalai (2010).

Artzner, Delbaen, Eber, and Heath (1999) define the concept of a coherent risk measure and showed that expected shortfall (*ES*) is coherent whereas *VaR* is not. Studying dynamic portfolio management based on *ES* and *VaR*, Basak and Shapiro (2001) found that when a large loss does occur, *ES* risk management leads to lower losses than *VaR* risk management. Cuoco, He, and Issaenko (2008) argued instead that *VaR* and *ES* risk management lead to equivalent results as long as the *VaR* and *ES* risk measures are recalculated often. Both Basak and Shapiro (2001) and Cuoco et al. (2008) assumed that returns are normally distributed. Chen (2008) and Taylor (2008) consider nonparametric *ES* methods.

For analyses of GARCH-based risk models more generally see Bali, Mo, and Tang (2008), Mancini and Trojani (2011), and Jalal and Rockinger (2008).

References

Artzner, P., Delbaen, F., Eber, J., Heath, D., 1999. Coherent measures of risk. Math. Finance 9, 203–228.

Bali, T., Mo, H., Tang, Y., 2008. The role of autoregressive conditional skewness and kurtosis in the estimation of conditional VaR. J. Bank. Finance 32, 269–282.

Basak, S., Shapiro, A., 2001. Value at risk based risk management: Optimal policies and asset prices. Rev. Financ. Stud. 14, 371–405.
Bollerslev, T., 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. Rev. Econ. Stat. 69, 542–547.
Brooks, C., Clare, A., Molle, J.D., Persand, G., 2005. A comparison of extreme value theory approaches for determining value at risk. J. Empir. Finance 12, 339–352.
Chen, S.X., 2008. Nonparametric estimation of expected shortfall. J. Financ. Econom 6, 87–107.
Christoffersen, P., Jacobs, K., Ornthanalai, C., 2010. Exploring time-varying jump intensities: Evidence from S&P 500 returns and options. Available from: SSRN, http://ssrn.com/abstract=1101733.
Cuoco, D., He, H., Issaenko, S., 2008. Optimal dynamic trading strategies with risk limits. Oper. Res. 56, 358–368.
Fernandez, C., Steel, M.F.J., 1998. On Bayesian modeling of fat tails and skewness. J Am. Stat. Assoc. 93, 359–371.
Giamouridis, D., 2006. Estimation risk in financial risk management: A correction. J. Risk 8, 121–125.
Hansen, B., 1994. Autoregressive conditional density estimation. Int. Econ. Rev. 35, 705–730.
Huisman, R., Koedijk, K., Kool, C., Palm, F., 2001. Tail-index estimates in small samples. J. Bus. Econ. Stat. 19, 208–216.
Jalal, A., Rockinger, M., 2008. Predicting tail-related risk measures: The consequences of using GARCH filters for Non-GARCH data. J. Empir. Finance 15, 868–877.
Jondeau, E., Rockinger, M., 2001. Gram-Charlier densities. J. Econ. Dyn. Control 25, 1457–1483.
Jondeau, E., Rockinger, M., 2003. Conditional volatility, skewness and kurtosis: Existence, persistence and comovements. J. Econ. Dyn. Control 27, 1699–1737.
Kuerster, K., Mittnik, S., Paolella, M., 2006. Value-at-Risk prediction: A comparison of alternative strategies. J. Financ. Econom. 4, 53–89.
Longin, F., 2000. From value at risk to stress testing: The extreme value approach. J. Bank. Finance 24, 1097–1130.
Longin, F., Solnik, B., 2001. Extreme correlation of international equity markets. J. Finance 56, 649–676.
Maheu, J., McCurdy, T., 2004. News arrival, jump dynamics and volatility components for individual stock returns. J. Finance 59, 755–794.
Mancini, L., Trojani, F., 2011. Robust value at risk prediction. J. Financ. Econom. 9, 281–313.
McNeil, A., 1997. Estimating the tails of loss severity distributions using extreme value theory. ASTIN Bull. 27, 117–137.
McNeil, A., 1999. Extreme value theory for risk managers. In: Internal modelling and CAD II. London: Risk Books, pp. 23–43.
McNeil, A., Frey, R., 2000. Estimation of tail-related risk measures for heteroskedastic financial time series: An extreme value approach. J. Empir. Finance 7, 271–300.
McNeil, A., Saladin, T., 1997. The peaks over thresholds method for estimating high quantiles of loss distributions. In: Proceedings of the 28th International ASTIN Colloquium, pp. 23–43, Cairns, Australia.
Ornthanalai, C., 2010. A new class of asset pricing models with Lévy processes: Theory and applications. Available from: SSRN, http://ssrn.com/abstract=1267432.
Poon, S.-H., Rockinger, M., Tawn, J., 2003. Extreme-value dependence measures and finance applications. Stat. Sin. 13, 929–953.
Taylor, J.W., 2008. Using exponentially weighted quantile regression to estimate value at risk and expected shortfall. J. Financ. Econom. 6, 382–406.

Empirical Exercises

Open the Chapter6Data.xlsx file from the companion site.

1. Construct a QQ plot of the S&P 500 returns divided by the unconditional standard deviation. Use the normal distribution. Compare your result with the top panel of Figure 6.2. (*Excel hint*: Use the NORMSINV function to calculate the standard normal quantiles.)
2. Copy and paste the estimated NGARCH(1,1) volatilities from Chapter 4.
3. Standardize the returns using the volatilities from exercise 2. Construct a QQ plot for the standardized returns using the normal distribution. Compare your result with the bottom panel of Figure 6.2.
4. Using QMLE, estimate the NGARCH(1,1)-$\tilde{t}(d)$ model. Fix the variance parameters at their values from exercise 3. Set the starting value of d equal to 10. (*Excel hint*: Use the GAMMALN function for the log-likelihood function of the standardized $t(d)$ distribution.)

   Construct a QQ plot for the standardized returns using the standardized $t(d)$ distribution. Compare your result with Figure 6.3. (*Excel hint*: Excel contains a two-sided quantile from the $t(d)$ distribution. To compute one-sided quantiles from the standardized $t(d)$ distribution, we use the relationship

$$\tilde{t}_p^{-1}(d) = \begin{cases} -\left|\text{tinv}(2p, d)\right| \sqrt{(d-2)/d}, & \text{if } p \leq 0.5 \\ \left|\text{tinv}(2(1-p), d)\right| \sqrt{(d-2)/d}, & \text{if } p > 0.5 \end{cases}$$

   where tinv is the function in Excel, and where $\tilde{t}_p^{-1}(d)$ is the standardized one-sided quantile we need for the QQ plot.)
5. Estimate the EVT model on the standardized portfolio returns using the Hill estimator. Use the 50 largest losses to estimate EVT. Calculate the 0.01% standardized return quantile implied by each of the following models: normal, $t(d)$, EVT, and Cornish-Fisher. Notice how different the 0.01% *VaR*s would be from these four models.
6. Construct the QQ plot using the EVT distribution for the 50 largest losses. Compare your result with Figure 6.7.
7. For each day in 2010, calculate the 1-day, 1% *VaR*s using the following methods: (a) RiskMetrics, that is, normal distribution with an exponential smoother on variance using the weight $\lambda = 0.94$; (b) NGARCH(1,1)-$\tilde{t}(d)$ with the parameters estimated in exercise 5; (c) Historical Simulation; and (d) Filtered Historical Simulation. Use a 251-day moving sample for Historical Simulation. Plot the *VaR*s.
8. Use the asymmetric t distribution to construct Figure 6.4.
9. Use the asymmetric t distribution to construct Figure 6.5.

The answers to these exercises can be found in the Chapter6Results.xlsx file, which is available from the companion site.

For more information see the companion site at
http://www.elsevierdirect.com/companions/9780123744487

7 Covariance and Correlation Models

1 Chapter Overview

Chapters 4 through 6 covered various aspects of modeling the aggregate portfolio return. The univariate methods in those chapters can also be used to model the return on each individual asset. Chapters 7 through 9 will cover multivariate risk models. They will enable us to join together the univariate asset level models in Chapters 4 through 6.

Although modeling the aggregate portfolio return directly is useful for passive portfolio risk measurement, it is not as useful for active risk management. In order to perform sensitivity analysis (for example, what happens to my portfolio risk if I buy another share of IBM?) and in order to assess the benefits of diversification, we need models of the dependence between the return on individual assets. We will proceed on this front in three steps. Chapter 7 (the present chapter) will model dynamic covariance and correlation, which together with the dynamic volatility models in Chapters 4 and 5 can be used to construct covariance matrices. Chapter 8 will develop some important simulation tools such as Monte Carlo and bootstrapping, which are needed for multiperiod risk assessments. Chapter 9 will introduce copula models, which can be used to link together the nonnormal univariate distributions in Chapter 6. Correlation models only allow for linear dependence between asset returns whereas copula models allow for nonlinear dependence.

The objective of this chapter is to model the linear dependence, or correlation, between returns on different assets, such as IBM and Microsoft stocks, or on different classes of assets, such as stock indices and foreign exchange (FX) rates. Once this is done, we will be able to calculate risk measures on portfolios of securities such as stocks, bonds, and foreign exchange rates for many different combinations of portfolio weights.

We first present a general model of portfolio risk for large-dimensional portfolios with many assets and consider ways to reduce the problem of dimensionality. Just as the main topic of the previous chapter was modeling the dynamic aspects of variance, the main topic of this chapter is modeling the dynamic aspects of correlation. We then consider dynamic correlation models of varying degrees of sophistication, both in terms of their specification and of the information required to calculate them.

Just as Chapters 4 and 5 temporarily assumed the univariate normal distribution in order to focus on volatility modeling, in this chapter we will assume the multivariate normal distribution for the purpose of covariance and correlation modeling. We hasten

Elements of Financial Risk Management. DOI: 10.1016/B978-0-12-374448-7.00007-5

to add that the assumption of multivariate normality is made for convenience and is not realistic. Important methods for dealing with the multivariate nonnormality evident in daily returns will be discussed in Chapters 8 and 9.

2 Portfolio Variance and Covariance

We now establish some notation that is necessary to study the risk of portfolios consisting of an arbitrary number of securities, n. The return on the portfolio on day $t+1$ is defined as

$$r_{PF,t+1} = \sum_{i=1}^{n} w_{i,t} r_{i,t+1}$$

where the sum is taken over the n securities in the portfolio. $w_{i,t}$ denotes the relative weight of security i at the end of day t. Note that as in Chapter 1, lowercase r denotes rates of return rather than log returns, R.

For daily log returns the portfolio return relationship will hold approximately

$$R_{PF,t+1} \approx \sum_{i=1}^{n} w_{i,t} R_{i,t+1}$$

Chapters 4 through 6 modeled the univariate $r_{PF,t+1}$ time series (or $R_{PF,t+1}$), Chapters 7 through 9 will model the multivariate time series $r_{i,t+1}, i = 1, 2, \ldots, n$ (or $R_{i,t+1}$). The models we will develop later for dynamic covariance and correlation can equally well be used for log returns and rates of returns, but $\sigma^2_{PF,t+1}$, *VaR*, and *ES* computations that depend on the portfolio weights, $w_{i,t}$, will only be exact when we use the rate of return definition, $r_{PF,t+1}$.

The variance of the portfolio can be written as

$$\sigma^2_{PF,t+1} = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,t} w_{j,t} \sigma_{ij,t+1} = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,t} w_{j,t} \sigma_{i,t+1} \sigma_{j,t+1} \rho_{ij,t+1}$$

where $\sigma_{ij,t+1}$ and $\rho_{ij,t+1}$ are respectively the covariance and correlation between security i and j on day $t+1$. Notice we have $\sigma_{ij,t+1} = \sigma_{ji,t+1}$ and $\rho_{ij,t+1} = \rho_{ji,t+1}$ for all i and j. We also have $\rho_{ii,t+1} = 1$ and $\sigma_{ii,t+1} = \sigma^2_{i,t+1}$ for all i.

Using vector notation, we will write

$$\sigma^2_{PF,t+1} = w_t' \Sigma_{t+1} w_t$$

where w_t is the $n \times 1$ vector of portfolio weights and Σ_{t+1} is the $n \times n$ covariance matrix of returns. In the case where $n = 2$, we simply have

$$\begin{aligned}
\sigma^2_{PF,t+1} &= \begin{bmatrix} w_{1,t} & w_{2,t} \end{bmatrix} \begin{bmatrix} \sigma^2_{1,t+1} & \sigma_{12,t+1} \\ \sigma_{12,t+1} & \sigma^2_{2,t+1} \end{bmatrix} \begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} \\
&= w^2_{1,t}\sigma^2_{1,t+1} + w^2_{2,t}\sigma^2_{2,t+1} + 2 w_{1,t} w_{2,t} \sigma_{12,t+1}
\end{aligned}$$

as $\sigma_{21,t+1} = \sigma_{12,t+1}$.

If we are willing to assume that returns are multivariate normal, then the portfolio return, which is just a linear combination of asset returns, will be normally distributed, and we have

$$VaR^p_{t+1} = -\sigma_{PF,t+1}\Phi^{-1}_p$$

and

$$ES^p_{t+1} = \sigma_{PF,t+1}\frac{\phi\left(\Phi^{-1}_p\right)}{p}$$

Notice that even if we have already constructed volatility forecasts for each of the securities in the portfolio, then we still have to model and forecast all the correlations. If we have n assets, then we will have $n(n-1)/2$ different correlations, so if n is 100, then we'll have 4950 correlations to model, which would be a daunting task. We will therefore explicitly be looking for methods that are able to handle large-dimensional portfolios.

2.1 Exposure Mappings

A very simple way to reduce the dimensionality of the portfolio variance is to impose a factor structure using observed market returns as factors. In the extreme case we may be able to assume that the portfolio return is just the (systematic) market return plus a portfolio specific (idiosyncratic) risk term as in

$$r_{PF,t+1} = r_{Mkt,t+1} + \varepsilon_{t+1}$$

where we assume that the idiosyncratic risk term, ε_{t+1}, is independent of the market return and has constant variance.

The portfolio variance in this case is

$$\sigma^2_{PF,t+1} = \sigma^2_{Mkt,t+1} + \sigma^2_\varepsilon$$

In the case of a very well diversified stock portfolio, for example, it may be reasonable to assume that the variance of the portfolio equals that of the S&P 500 market index. In this case, only one volatility—that of the S&P 500 index return—needs to be modeled, and no correlation modeling is necessary. This is referred to as index mapping and can be written as

$$\sigma^2_{PF,t+1} \approx \sigma^2_{Mkt,t+1}$$

The 1-day *VaR* assuming normality is simply

$$VaR^p_{t+1} = -\sigma_{Mkt,t+1}\Phi^{-1}_p$$

More generally, in portfolios that contain systematic risk, we have

$$r_{PF,t+1} = \beta r_{Mkt,t+1} + \varepsilon_{t+1}$$

so that

$$\sigma^2_{PF,t+1} = \beta^2 \sigma^2_{Mkt,t+1} + \sigma^2_{\varepsilon}$$

If the portfolio is well diversified so that systematic market risk explains a large part of the variation in the portfolio return then the portfolio-specific idiosyncratic risk can be ignored, and we can pose a linear relationship between the portfolio and the market index and use the so-called beta mapping, as in

$$\sigma^2_{PF,t+1} \approx \beta^2 \sigma^2_{Mkt,t+1}$$

In this case, only an estimate of β is necessary and no further correlation modeling is needed.

Finally, the risk manager of a large-scale portfolio may consider risk as mainly coming from a reasonable number of factors n_F where $n_F << n$ so that we have many fewer risk factors than assets. The exact choice of factors depends highly on the particular portfolio at hand, but they could be, for example, country equity indices, FX rates, or commodity price indices. Let us assume that we need 10 factors. We can write the 10-factor return model as

$$r_{PF,t+1} = \beta_1 r_{F1,t+1} + \cdots + \beta_{10} r_{F10,t+1} + \varepsilon_{t+1}$$

where again the ε_{t+1} are assumed to be independent of the risk factors.

In this case, it makes sense to model the variances and correlations of these risk factors and assign exposures to each factor to get the portfolio variance. The portfolio variance in this general factor structure can be written

$$\sigma^2_{PF,t+1} = \beta_F' \Sigma^F_{t+1} \beta_F + \sigma^2_{\varepsilon}$$

where $w_{F,t}$ is a vector of exposures to each risk factor and where Σ^F_{t+1} is the covariance matrix of the returns from the risk factors. Again, if the factor model explains a large part of the portfolio return variation, then we can assume that

$$\sigma^2_{PF,t+1} \approx w_{F,t}' \Sigma^F_{t+1} w_{F,t}$$

2.2 GARCH Conditional Covariances

Suppose the portfolio under consideration contains n assets. Alternatively, we can think of the risk manager as having chosen n_F risk factors to be the main drivers of the risk in the portfolio. In either case, a covariance matrix must be estimated. To simplify notation, let us assume that we need an n-dimensional covariance matrix where n is 10 or larger. In portfolios with relatively few assets n will be the number of assets and in portfolios with many assets n will be the number of risk factors. We will refer to the assets or risk factors generically as “assets” in the following.

We now turn to various methods for constructing the time-varying covariance matrix Σ_{t+1}. Arguably the simplest way to model time-varying covariances is to rely

on plain rolling averages, a method that we considered for volatility in Chapter 4. For the covariance between asset (or risk factor) i and j, we can simply estimate

$$\sigma_{ij,t+1} = \frac{1}{m}\sum_{\tau=1}^{m} R_{i,t+1-\tau}R_{j,t+1-\tau}$$

where m is the number of days used in the moving estimation window. This estimate is very easy to construct but it is not satisfactory due to the dependence on the choice of m and the equal weighting put on past cross products of returns. Notice that, as in previous chapters, we assume the average expected return on each asset (or risk factor) is simply zero. Figure 7.1 shows the rolling covariance between the S&P 500 and the return on an index of 10-year treasury notes when $m = 25$.

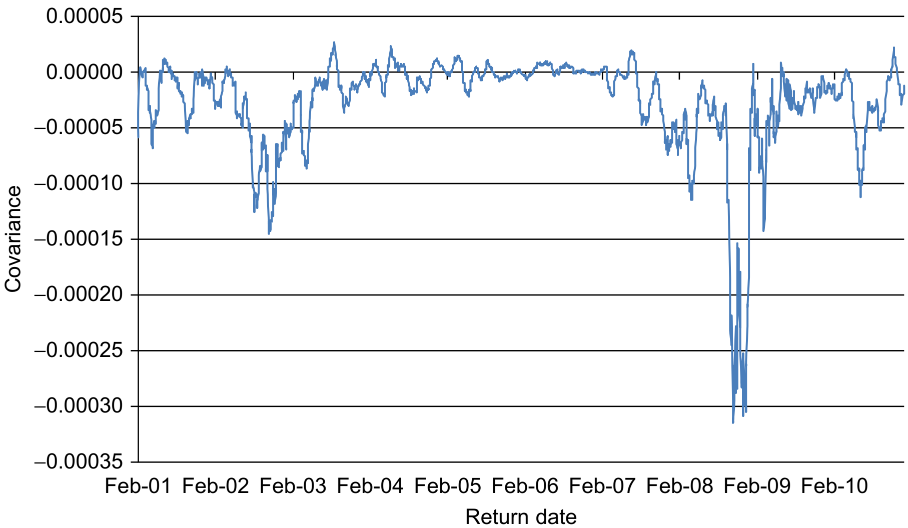
In order to avoid equal weighting we can instead use a simple exponential smoother model on the covariances, and let

$$\sigma_{ij,t+1} = (1-\lambda)R_{i,t}R_{j,t} + \lambda\sigma_{ij,t}$$

where $\lambda = 0.94$ as it was for the corresponding volatility model in the previous chapters. Figure 7.2 shows the exponential smoother covariance between the S&P 500 and the 10-year treasury note.

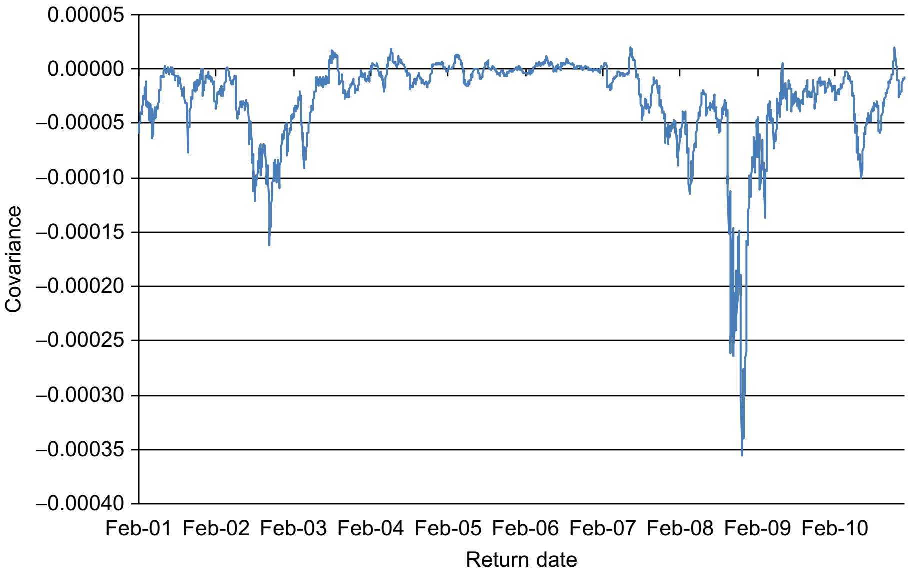
The caveats that applied to the exponential smoother volatility model in Chapter 4 apply to the exponential smoother covariance model as well. The restriction that the coefficient $(1-\lambda)$ on the cross product of returns $\left(R_{i,t}R_{j,t}\right)$ and the coefficient λ on the

Figure 7.1 Rolling covariance between S&P 500 and 10-year treasury note index.



Notes: We plot the rolling covariance computed on a moving window of 25 trading days.

Figure 7.2 Exponentially smoothed covariance between S&P 500 and 10-year treasury note index.



Notes: We use the RiskMetrics exponential smoother with $\lambda = 0.94$ on the cross product of returns.

past covariance $(\sigma_{ij,t})$ sum to one is not necessarily desirable. It implies that there is no mean-reversion in covariance. Consider two highly correlated assets. If the forecast for tomorrow's covariance happens to be low today, then the forecast will be low for all future horizons in the exponential smoother model. The forecast from the model will not revert back to its (higher) mean when the horizon increases.

We can instead consider models with mean-reversion in covariance. For example, a GARCH-style specification for covariance would be

$$\sigma_{ij,t+1} = \omega_{ij} + \alpha R_{i,t} R_{j,t} + \beta \sigma_{ij,t}$$

which will tend to revert to its long-run average covariance, which equals

$$\sigma_{ij} = \omega_{ij}/(1 - \alpha - \beta)$$

Notice that so far we have not allowed the persistence parameters λ in RiskMetrics and α and β in GARCH to vary across pairs of securities in the covariance models. This is no coincidence. It must be done to guarantee that the portfolio variance will be positive regardless of the portfolio weights, w_t. We will say that a covariance matrix Σ_{t+1} is internally consistent if for all possible vectors w_t of portfolio weights we have

$$w_t' \Sigma_{t+1} w_t \geq 0$$

This corresponds to saying that the covariance matrix is positive semidefinite. It is ensured by estimating volatilities and covariances in an internally consistent fashion. For example, relying on exponential smoothing using the same λ for every volatility and every covariance will work. Similarly, using a GARCH(1,1) model with α and β identical across variances and covariances and with long-run variances and covariances estimated consistently will work as well.

Unfortunately, it is not clear that the persistence parameters λ, α, and β should be the same for all variances and covariance. We therefore next consider methods that are not subject to this restriction.

3 Dynamic Conditional Correlation (DCC)

We now turn to the modeling of correlation rather than covariance. This is motivated by the desire to free up the restriction that variances and covariance have the same persistence parameters. We also want to assess if the time-variation in covariances arises solely from time-variation in the volatilities or if correlation has its own dynamic pattern. There is ample empirical evidence that correlations increase during financial turmoil and thereby increase risk even further; therefore, modeling correlation dynamics is crucial to a risk manager.

From Chapter 3, correlation is defined from covariance and volatility by

$$\rho_{ij,t+1} = \sigma_{ij,t+1}/(\sigma_{i,t+1}\sigma_{j,t+1})$$

If, for example, we have the RiskMetrics model, then

$$\sigma_{ij,t+1} = (1-\lambda)R_{i,t}R_{j,t} + \lambda\sigma_{ij,t}, \quad \text{for all } i,j$$

and then we get the implied dynamic correlations

$$\rho_{ij,t+1} = \frac{(1-\lambda)R_{i,t}R_{j,t} + \lambda\sigma_{ij,t}}{\sqrt{((1-\lambda)R_{i,t}^2 + \lambda\sigma_{i,t}^2)((1-\lambda)R_{j,t}^2 + \lambda\sigma_{j,t}^2)}}$$

which, of course, is not particularly intuitive. We therefore now consider models where the dynamic correlation is modeled directly.

The definition of correlation can be rearranged to provide the decomposition of covariance into volatility and correlation

$$\sigma_{ij,t+1} = \sigma_{i,t+1}\sigma_{j,t+1}\rho_{ij,t+1}$$

In matrix notation, we can write

$$\Sigma_{t+1} = D_{t+1}\Upsilon_{t+1}D_{t+1}$$

where D_{t+1} is a matrix of standard deviations, $\sigma_{i,t+1}$, on the ith diagonal and zero everywhere else, and where Υ_{t+1} is a matrix of correlations, $\rho_{ij,t+1}$, with ones on the diagonal. In the two-asset case, we have

$$\Sigma_{t+1} = \begin{bmatrix} \sigma^2_{1,t+1} & \sigma_{12,t+1} \\ \sigma_{12,t+1} & \sigma^2_{2,t+1} \end{bmatrix} = \begin{bmatrix} \sigma_{1,t+1} & 0 \\ 0 & \sigma_{2,t+1} \end{bmatrix} \begin{bmatrix} 1 & \rho_{12,t+1} \\ \rho_{12,t+1} & 1 \end{bmatrix} \begin{bmatrix} \sigma_{1,t+1} & 0 \\ 0 & \sigma_{2,t+1} \end{bmatrix}$$

We will consider the volatilities of each asset to already have been estimated through GARCH or one of the other methods considered in Chapter 4 or 5. We can then standardize each return by its dynamic standard deviation to get the standardized returns,

$$z_{i,t+1} = R_{i,t+1}/\sigma_{i,t+1} \text{ for all } i$$

By dividing the returns by their conditional standard deviation, we create variables, $z_{i,t+1}$, $i = 1, 2, \ldots, n$, which all have a conditional standard deviation of one. The conditional covariance of the $z_{i,t+1}$ variables equals the conditional correlation of the raw returns as can be seen from

$$\begin{aligned} E_t\left(z_{i,t+1} z_{j,t+1}\right) &= E_t\left(\left(R_{i,t+1}/\sigma_{i,t+1}\right)\left(R_{j,t+1}/\sigma_{j,t+1}\right)\right) \\ &= E_t\left(R_{i,t+1} R_{j,t+1}\right)/(\sigma_{i,t+1}\sigma_{j,t+1}) \\ &= \sigma_{ij,t+1}/(\sigma_{i,t+1}\sigma_{j,t+1}) \\ &= \rho_{ij,t+1}, \text{ for all } i, j \end{aligned}$$

Thus, modeling the conditional correlation of the raw returns is equivalent to modeling the conditional covariance of the standardized returns.

3.1 Exponential Smoother Correlations

We first consider simple exponential smoothing correlation models. Let the correlation dynamics be driven by $q_{ij,t+1}$, which gets updated by the cross product of the standardized returns $z_{i,t}$ and $z_{j,t}$ as in

$$q_{ij,t+1} = (1 - \lambda)\left(z_{i,t} z_{j,t}\right) + \lambda q_{ij,t}, \quad \text{for all } i, j$$

The exact conditional correlation can now be obtained by normalizing the $q_{ij,t+1}$ variable as in

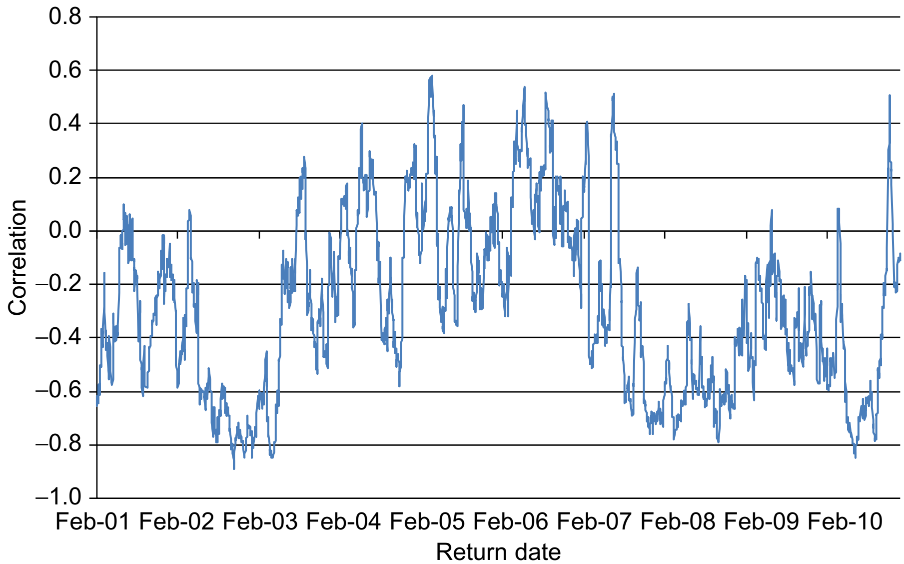
$$\rho_{ij,t+1} = \frac{q_{ij,t+1}}{\sqrt{q_{ii,t+1} q_{jj,t+1}}}$$

The reason we need to do the normalization is to ensure that

$$-1 < \rho_{ij,t+1} < +1$$

on each day.

Figure 7.3 Exponentially smoothed correlation between S&P 500 and 10-year treasury note index.



Notes: We use the DCC model with exponential smoother dynamics to plot the dynamic correlation. The individual variances are modeled using NGARCH.

Figure 7.3 shows the exponential smoothed DCC correlations for the S&P 500 and 10-year treasury note example. Notice the dramatic changes in correlation over time.

The DCC model requires estimation of λ and we will discuss estimation in detail later.

3.2 Mean-Reverting Correlation

Just as we did for volatility and covariance models, we may want to consider a generalization of the exponential smoothing correlation model, which allows for correlations to revert to a long-run average correlation, $\rho_{ij} = E\left[z_{i,t}z_{j,t}\right]$. We can consider GARCH(1,1)-type specifications of the form

$$q_{ij,t+1} = \rho_{ij} + \alpha\left(z_{i,t}z_{j,t} - \rho_{ij}\right) + \beta\left(q_{ij,t} - \rho_{ij}\right)$$

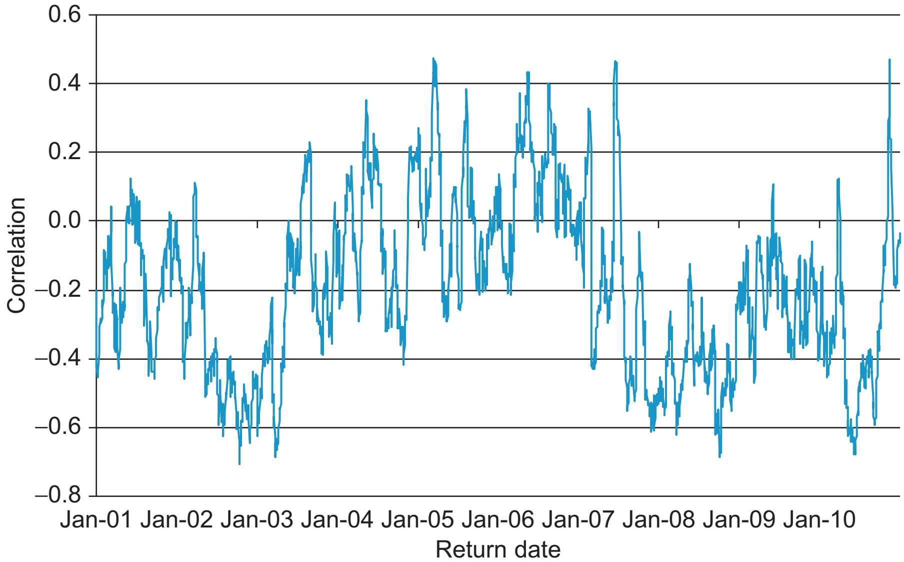
If we rely on correlation targeting, and set $\bar{\rho}_{ij} = \frac{1}{T}\sum_{t=1}^{T} z_{i,t}z_{j,t}$, then we have

$$q_{ij,t+1} = \bar{\rho}_{ij} + \alpha\left(z_{i,t}z_{j,t} - \bar{\rho}_{ij}\right) + \beta\left(q_{ij,t} - \bar{\rho}_{ij}\right)$$

Again we have to normalize to get the conditional correlations

$$\rho_{ij,t+1} = \frac{q_{ij,t+1}}{\sqrt{q_{ii,t+1}q_{jj,t+1}}}$$

Figure 7.4 Mean-reverting correlation between S&P 500 and 10-year treasury note.



Notes: We use the DCC model with mean-reverting dynamics to plot the dynamic correlation. The individual variances are modeled using NGARCH.

The key thing to note about this model is that the correlation persistence parameters α and β are common across i and j. Thus, the model implies that the persistence of the correlation between any two assets in the portfolio is the same. It does not, however, imply that the level of the correlations at any time are the same across pairs of assets. The level of correlation is controlled by ρ_{ij} and will thus vary over i and j. It does also not imply that the persistence in correlation is the same as the persistence in volatility. The persistence in volatility can vary from asset to asset, and it can vary from the persistence in correlation between the assets. But the model does imply that the persistence in correlation is constant across assets. Figure 7.4 shows the GARCH(1,1) correlations for the S&P 500 and 10-year treasury note example.

We can write the DCC models in matrix notation as

$$Q_{t+1} = (1-\lambda)\left(z_t z_t'\right) + \lambda Q_t$$

for the exponential smoother, and for the mean-reverting DCC, we can write

$$Q_{t+1} = E\left[z_t z_t'\right](1-\alpha-\beta) + \alpha\left(z_t z_t'\right) + \beta Q_t$$

In the two-asset case for the mean-reverting model, we have

$$\begin{aligned} Q_{t+1} &= \begin{bmatrix} q_{11,t+1} & q_{12,t+1} \\ q_{12,t+1} & q_{22,t+1} \end{bmatrix} \\ &= \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix}(1-\alpha-\beta) + \alpha \begin{bmatrix} z_{1,t}^2 & z_{1,t}z_{2,t} \\ z_{1,t}z_{2,t} & z_{2,t}^2 \end{bmatrix} + \beta \begin{bmatrix} q_{11,t} & q_{12,t} \\ q_{12,t} & q_{22,t} \end{bmatrix} \end{aligned}$$

where ρ_{12} is the unconditional correlation between the two assets, which can be estimated in advance as

$$\bar{\rho}_{12} = \frac{1}{T}\sum_{t=1}^{T} z_{1,t}z_{2,t}$$

An important feature of these models is that the matrix Q_{t+1} is positive semidefinite since it is a weighted average of positive semidefinite and positive definite matrices. This will in turn ensure that the correlation matrix Υ_{t+1} and the covariance matrix Σ_{t+1} will be positive semidefinite as required.

Another important practical advantage of the DCC model is that we can estimate the parameters in a sequential fashion. First, all the individual variances are estimated one by one using one of the methods from Chapter 4 or 5. Second, the returns are standardized and the unconditional correlation matrix is estimated. Third, the correlation persistence parameters α and β are estimated. The key issue is that only very few parameters are estimated simultaneously using numerical optimization. This feature makes the dynamic correlation models considered here extremely tractable for risk management of large portfolios. We now turn to the details of the estimation procedure.

3.3 Bivariate Quasi Maximum Likelihood Estimation

Fortunately, in estimating the dynamic conditional correlation models suggested earlier, we can rely on the quasi maximum likelihood estimation (QMLE) method, which we used for estimating the GARCH volatility models in Chapter 4.

Although a key benefit of the correlation models suggested here is that they are easy to estimate even for large portfolios, we will begin by analyzing the case of a portfolio consisting of only two assets. In this case, we can use the bivariate normal distribution function for $z_{1,t}$ and $z_{2,t}$ to write the log likelihood as

$$\ln\left(L_{c,12}\right) = -\frac{1}{2}\sum_{t=1}^{T}\left(\ln\left(1-\rho_{12,t}^2\right) + \frac{\left(z_{1,t}^2 + z_{2,t}^2 - 2\rho_{12,t}z_{1,t}z_{2,t}\right)}{\left(1-\rho_{12,t}^2\right)}\right)$$

where $\rho_{12,t}$ is given from the particular correlation model being estimated and the normalization rule. Note that we have omitted the constant term involving π, which is irrelevant when optimizing. In the simple exponential smoother example,

$$\rho_{12,t} = \frac{q_{12,t}}{\sqrt{q_{11,t}q_{22,t}}}$$

where

$$q_{11,t} = (1-\lambda)\left(z_{1,t-1}^2\right) + \lambda q_{11,t-1}$$
$$q_{12,t} = (1-\lambda)\left(z_{1,t-1}z_{2,t-1}\right) + \lambda q_{12,t-1}$$
$$q_{22,t} = (1-\lambda)\left(z_{2,t-1}^2\right) + \lambda q_{22,t-1}$$

We find the optimal correlation parameter(s), in this case λ, by maximizing the correlation log-likelihood function, $\ln(L_{c,12})$. To initialize the dynamics, we set $q_{11,0} = 1, q_{22,0} = 1$, and $q_{12,0} = \frac{1}{T}\sum_{t=1}^{T} z_{1,t}z_{2,t}$.

Notice that the variables that enter the likelihood are the standardized returns, z_t, and not the original raw returns, R_t themselves. We are essentially treating the standardized returns as actual observations here.

As before, the QMLE method will give us consistent but inefficient estimates. In theory, we could obtain more precise results by estimating all the volatility models and the correlation model simultaneously. In practice, this is not feasible for large portfolios. In realistic situations, we are forced to rely on a stepwise QMLE method where we first estimate the volatility model for each of the assets and second estimate the correlation models. This approach gives decent parameter estimates while avoiding numerical optimization in high dimensions.

In the case of the mean-reverting GARCH correlations we have the same likelihood function and correlation definition but now

$$
\begin{aligned}
q_{11,t+1} &= 1 + \alpha\left(z_{1,t}^2 - 1\right) + \beta\left(q_{11,t} - 1\right) \\
q_{12,t+1} &= \bar{\rho}_{12} + \alpha\left(z_{1,t}z_{2,t} - \bar{\rho}_{12}\right) + \beta\left(q_{12,t} - \bar{\rho}_{12}\right) \\
q_{22,t+1} &= 1 + \alpha\left(z_{2,t}^2 - 1\right) + \beta\left(q_{22,t} - 1\right)
\end{aligned}
$$

where $\bar{\rho}_{12}$ can be estimated using

$$\bar{\rho}_{12} = \frac{1}{T}\sum_{t=1}^{T} z_{1,t}z_{2,t}$$

Therefore we only have to find α and β using numerical optimization. Again, in order to initialize the dynamics, we set $q_{11,0} = 1, q_{22,0} = 1$, and $q_{12,0} = \bar{\rho}_{12} = \frac{1}{T}\sum_{t=1}^{T} z_{1,t}z_{2,t}$.

3.4 Composite Likelihood Estimation in Large Systems

In the general case of n assets in the portfolio, we typically have to rely on the $n-$dimensional normal distribution function to write the log likelihood as

$$\ln(L_c) = -\frac{1}{2}\sum_t \left(\log|\Upsilon_t| + z_t'\Upsilon_t^{-1}z_t\right)$$

where $|\Upsilon_t|$ denotes the determinant of the correlation matrix, Υ_t. Maximizing this likelihood can be very cumbersome if n is large. The correlation matrix Υ_t must be inverted on each day and for many possible values of the parameters in the model when doing the numerical search for optimal parameter values. When n is large the inversion of Υ_t will be slow and potentially inaccurate causing biases in the estimated parameter values.

Fortunately a very simple solution to this dimensionality problem is available in the DCC model. Rather than maximizing the n-dimensional log likelihood we can maximize the sum of the bivariate likelihoods

$$\ln(CL_c) = -\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j>i}\left(\ln\left(1-\rho_{ij,t}^2\right) + \frac{\left(z_{i,t}^2 + z_{j,t}^2 - 2\rho_{ij,t}z_{i,t}z_{j,t}\right)}{\left(1-\rho_{ij,t}^2\right)}\right)$$

Computationally this composite likelihood function is much easier to maximize than the likelihood function where the n-dimensional correlation matrix must be inverted numerically. Note that rather than using all the available bivariate likelihoods in $\ln(CL_c)$ we could just use a subset of them. The more we use the more precise our estimates will get.

3.5 An Asymmetric Correlation Model

So far we have considered only symmetric correlation models where the effect of two positive shocks is the same as the effect of two negative shocks of the same magnitude. But, just as we modeled the asymmetry in volatility (the leverage effect), we may want to allow for a down-market effect in correlation. This can be achieved using the asymmetric DCC model where

$$Q_{t+1} = (1-\alpha-\beta)E\left[z_t z_t'\right] + \alpha\left(z_t z_t'\right) + \beta Q_t + \gamma\left(\eta_t\eta_t' - E\left[\eta_t\eta_t'\right]\right)$$

where the $\eta_{i,t}$ for asset i is defined as the negative part of $z_{i,t}$ as follows:

$$\eta_{i,t} = \begin{cases} z_{i,t}, & \text{if } z_{i,t} < 0 \\ 0, & \text{if } z_{i,t} > 0 \end{cases} \quad \text{for all } i$$

Note that γ corresponds to a leverage effect in correlation: When γ is positive then the correlation for asset i and j will increase more when $z_{i,t}$ and $z_{j,t}$ are negative than in any other case. If we envision a scatterplot of $z_{i,t}$ and $z_{j,t}$, then $\gamma > 0$ will provide an extra increase in correlation when we observe an observation in the lower-left quadrant of the scatterplot. This captures a phenomenon often observed in markets for risky assets: Their correlation increases more in down markets ($z_{i,t}$ and $z_{j,t}$ both negative) than in up markets ($z_{i,t}$ and $z_{j,t}$ both positive). The basic DCC model does not directly capture this down-market effect but the asymmetric DCC model does.

4 Estimating Daily Covariance from Intraday Data

In Chapter 5 we considered methods for daily volatility estimation and forecasting that made use of intraday data. These methods can be extended to covariance estimation as well. When constructing RVs in Chapter 5 our biggest concern was biases arising from illiquidity effects: bid–ask bounces for example would bias upward our RV measure if we constructed intraday returns at a frequency that is too high.

The main concern when computing realized covariance is not bid–ask bounces but rather the asynchronicity of intraday prices across assets. Asynchronicity of intraday prices will cause a bias toward zero of realized covariance unless we estimate it carefully. Because asset covariances are typically positive a bias toward zero means we will be underestimating covariance and thus underestimating portfolio risk. This is clearly not a mistake we want to make.

4.1 Realized Covariance

Consider first daily covariance estimation using, say, 1-minute returns. As in Chapter 5, let the jth observation on day $t+1$ for asset 1 be denoted $S_{1,t+j/m}$. Then the jth return on day $t+1$ is

$$R_{1,t+j/m} = \ln(S_{1,t+j/m}) - \ln(S_{1,t+(j-1)/m}), \quad \text{for } j = 1, 2, \ldots, m$$

Observing m returns within a day for two assets recorded at exactly the same time intervals, we can in principle calculate an estimate of the realized daily covariance from the intraday cross product of returns simply as

$$RCov^{m}_{12,t+1} = \sum_{j=1}^{m} R_{1,t+j/m} R_{2,t+j/m}$$

Given estimates of the two volatilities, the realized correlation can, of course, then easily be calculated as

$$\rho^{m}_{12,t+1} = RCov^{m}_{12,t+1} / \sqrt{RV^{m}_{1,t+1} RV^{m}_{2,t+1}}$$

where $RV^{m}_{1,t+1}$ is the All RV estimator defined in Chapter 5 computed for asset 1.

However, from Chapter 5 we quickly realize that using the All RV estimate based on all m intraday returns is not a good idea because of the biases arising from illiquidity at high frequencies. We can instead rely on the Average (Avr) RV estimator, which averages across a number of sparse (using lower-frequency returns) RVs. Using the averaging idea for the RCov as well we would then have

$$RCorr^{Avr}_{12,t+1} = RCov^{Avr}_{12,t+1} / \sqrt{RV^{Avr}_{1,t+1} RV^{Avr}_{2,t+1}}$$

where RV^{Avr} is as defined in Chapter 5 and where $RCov^{Avr}$ can also be computed as the average of, say, 15 sparse $RCov^{s}_{12,t+1}$ estimators computed on overlapping 15-minute grids.

Going from All RV to Average RV will fix the bias problems in the RV estimates but it will unfortunately not fix the bias in the RCov estimates: Asynchronicity will still cause a bias toward zero in RCov.

The current best practice for alleviating the asynchronicity bias in daily RCov relies on changing the time scale of the intraday observations. When we observe intraday

prices on n assets the prices all arrive randomly throughout the day and randomly across assets. The trick for dealing with asynchronous data is to synchronize them using so-called refresh times.

Let $\tau(1)$ be the first time point on day $t+1$ when all assets have changed their price at least once since market open. Let $\tau(2)$ be the first time point on day $t+1$ when all assets have changed their price at least once since $\tau(1)$, and so on for $\tau(j)$, $j = 1, 2, \ldots, N$. The synchronized intraday returns for the n assets can now be computed using the $\tau(j)$ time points. For assets 1 and 2 we have

$$R_{1,\tau(j)} = \ln(S_{1,\tau(j)}) - \ln(S_{1,\tau(j-1)}), \quad \text{for } j = 1, 2, \ldots, N$$
$$R_{2,\tau(j)} = \ln(S_{2,\tau(j)}) - \ln(S_{2,\tau(j-1)}), \quad \text{for } j = 1, 2, \ldots, N$$

so that we can define the synchronized realized covariance between them as

$$RCov_{12,t+1}^{Sync} \equiv \sum_{j=1}^{N} R_{1,\tau(j)} R_{2,\tau(j)}$$

If realized variances are computed from the same refresh grid of prices

$$RV_{1,t+1}^{Sync} \equiv \sum_{j=1}^{N} R_{1,\tau(j)}^2, \quad \text{and} \quad RV_{2,t+1}^{Sync} \equiv \sum_{j=1}^{N} R_{2,\tau(j)}^2$$

then the variance-covariance matrix

$$\begin{bmatrix} RV_{1,t+1}^{Sync} & RCov_{12,t+1}^{Sync} \\ RCov_{12,t+1}^{Sync} & RV_{2,t+1}^{Sync} \end{bmatrix}$$

will be positive definite.

The synchronized RV and RCov estimates can be further refined by correcting for autocorrelation in the cross products of intraday returns as we did for RV in Chapter 5.

The RCov and RV measures can be used to build multivariate models for forecasting covariance and correlation. Some of the relevant references are listed at the end of the chapter.

4.2 Range-Based Covariance Using No-Arbitrage Conditions

Aside from the important synchronization problems, it is relatively straightforward to generalize the idea of realized volatility to realized correlation. However, extending range-based volatility to range-based correlation is not obvious because the cross product of the ranges does not capture covariance.

However, sometimes asset prices are linked together by no-arbitrage restrictions, and if so then range-based covariance can be constructed. Consider, for example, the case where S_1 is the US\$/yen FX rate, and S_2 is the Euro/US\$ FX rate. If we define S_3

to be the Euro/yen FX rate, then by ruling out arbitrage opportunities, we can write

$$S_{3,t} = S_{1,t}S_{2,t}$$
$$S_{3,t+1} = S_{1,t+1}S_{2,t+1}$$

Therefore, the log returns can be written

$$R_{3,t+1} = R_{1,t+1} + R_{2,t+1}$$

and the variances as

$$\sigma^2_{3,t+1} = \sigma^2_{1,t+1} + \sigma^2_{2,t+1} + 2\sigma_{12,t+1}$$

Thus, we can rearrange to get the covariance between US$/yen and Euro/US$ from

$$\sigma_{12,t+1} = \left(\sigma^2_{3,t+1} - \sigma^2_{1,t+1} - \sigma^2_{2,t+1}\right)/2$$

If we then use one of the range-based proxies from Chapter 5, for example

$$RP_{i,t+1} \approx 0.361D^2_{i,t+1} = .361\left[\ln\left(S^{High}_{i,t+1}\right) - \ln\left(S^{Low}_{i,t+1}\right)\right]^2, \quad \text{for } i = 1,2,3$$

we can define the range-based covariance proxy

$$\begin{aligned} RPCov_{12,t+1} &\equiv \left(RP_{3,t+1} - RP_{1,t+1} - RP_{2,t+1}\right)/2 \\ &\approx 0.185\left(D^2_{3,t+1} - D^2_{1,t+1} - D^2_{2,t+1}\right) \end{aligned}$$

Similar arbitrage arguments can be made between spot and futures prices and between portfolios and individual assets assuming of course that the range prices can be found on all the involved series.

Finally, as we suggested for volatility in Chapter 5, range-based proxies for covariance can be used as regressors in GARCH covariance models. Consider, for example,

$$\sigma_{ij,t+1} = \omega_{ij} + \alpha R_{i,t}R_{j,t} + \beta\sigma_{ij,t} + \gamma_{RP}RP_{ij,t+1}$$

Including the range-based covariance estimate in a GARCH model instead of using it by itself will have the beneficial effect of smoothing out some of the inherent noise in the range-based estimate of covariance.

5 Summary

Risk managers who want to calculate risk measures such as Value-at-Risk and Expected Shortfall for different portfolio allocations need to construct the matrix of variances and covariances for potentially large sets of assets. If returns are assumed to be normally distributed with a mean of zero, then the covariance matrix is all that

is needed to calculate the *VaR*. This chapter thus considered methods for constructing the covariance matrix. First, we presented simple rolling estimates of covariance, followed by simple exponential smoothing and GARCH models of covariance. We then discussed the important issue of estimating variances and covariances in an internally consistent way so as to ensure that the covariance matrix is positive semidefinite and therefore generates sensible portfolio variances for all possible portfolio weights. This discussion led us to consider modeling the conditional correlation rather than the conditional covariance. We presented a simple framework for dynamic correlation modeling, which is based on standardized returns and which thus relies on preestimated volatility models such as those discussed in Chapters 4 and 5. Finally, methods for daily covariance and correlation estimation that make use of intraday information were introduced.

Further Resources

The choice of risk factors may be obvious for some portfolios, but in general it is not. It is therefore useful to let the return data help when deciding on what the factors should look like and how many factors we need. The choice of factors in a variety of portfolios is discussed in detail in Connor et al. (2010). A nice overview of the mechanics of assigning risk factor exposures can be found in Jorion (2006).

Bollerslev et al. (1988), Bollerslev (1990), Bollerslev and Engle (1993), and Engle and Kroner (1995) are some classic references on the first generation of multivariate GARCH models. See also the recent survey in Bauwens et al. (2006).

The conditional correlation model in this chapter is developed in Engle (2002), Engle and Sheppard (2001), and Tse and Tsui (2002). Aielli (2009) derives a refinement to the QMLE DCC estimation procedure described in this chapter. Cappiello et al. (2006) and Hafner and Franses (2009) develop extensions and alternatives to the basic DCC model. Composite likelihood estimation of DCC models is suggested in Engle et al. (2009). For a large-scale application of DCC models to international equity markets see Christoffersen et al. (2011).

Asynchronicity in returns is not just an issue in intraday data. It can also be a problem in daily returns for illiquid assets or for assets from markets that close at different times of the day. Burns et al. (1998) and Audrino and Buhlmann (2004) develop vector ARMA methods to deal with biases in correlation. Scholes and Williams (1977) use measurement error models to analyze bias of the beta estimate in the market model when daily closing prices are stale.

The construction of realized covariances is detailed in Barndorff-Nielsen et al. (2011), which also contains useful information on the cleaning of intraday data. Forecasting models using realized covariance and correlation are built in Bauer and Vorkink (2011), Chiriac and Voev (2011), Hansen et al. (2010), Jin and Maheu (2010), Voev (2008), and Noureldin et al. (2011).

Range-based covariance estimation is considered in Brandt and Diebold (2006), who also discuss ways to ensure positive semidefiniteness of the covariance matrix. Foreign exchange covariances estimated from intraday returns are reported in Andersen et al. (2001).

Finally, methods for the evaluation of covariance and correlation forecasts can be found in Patton and Sheppard (2009).

References

Aielli, G., 2009. Dynamic conditional correlations: On properties and estimation. Available from: SSRN, http://ssrn.com/abstract=1507743.

Andersen, T., Bollerslev, T., Diebold, F., 2001. The distribution of realized exchange rate volatility. J. Am. Stat. Assoc. 96, 42–55.

Audrino, F., Buhlmann, P., 2004. Synchronizing multivariate financial time series. J. Risk 6, 81–106.

Barndorff-Nielsen, O., Hansen, P., Lunde, A., Shephard, N., 2011. Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. J. Econom. forthcoming.

Bauer, G.H., Vorkink, K., 2011. Forecasting multivariate realized stock market volatility. J. Econom. 160, 93–101.

Bauwens, L., Laurent, S., Rombouts, J., 2006. Multivariate GARCH models: A survey. J. Appl. Econom. 21, 79–109.

Bollerslev, T., 1990. Modeling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. Rev. Econ. Stat. 72, 498–505.

Bollerslev, T., Engle, R., 1993. Common persistence in conditional variances. Econometrica 61, 167–186.

Bollerslev, T., Engle, R., Wooldridge, J., 1988. A capital asset pricing model with time varying covariances. J. Polit. Econ. 96, 116–131.

Brandt, M., Diebold, F., 2006. A no-arbitrage approach to range-based estimation of return covariances and correlations. J. Bus. 79, 61–74.

Burns, P., Engle, R., Mezrich, J., 1998. Correlations and volatilities of asynchronous data. J. Derivatives 5, 7–18.

Cappiello, L., Engle, R., Sheppard, K., 2006. Asymmetric dynamics in the correlations of global equity and bond returns. J. Financ. Econom. 4, 537–572.

Chiriac, R., Voev, V., 2011. Modelling and forecasting multivariate realized volatility. J. Appl. Econom. forthcoming.

Christoffersen, P., Errunza, V., Jacobs, K., Langlois, H., 2011. Is the potential for international diversification disappearing? Available from: SSRN, http://ssrn.com/abstract=1573345.

Connor, G., Goldberg, L., Korajczyk, R., 2010. Portfolio Risk Analysis. Princeton University Press, Princeton, NJ.

Engle, R., 2002. Dynamic conditional correlation: A simple class of multivariate GARCH models. J. Bus. Econ. Stat. 20, 339–350.

Engle, R., Kroner, F., 1995. Multivariate simultaneous generalized ARCH. Econom. Theory 11, 122–150.

Engle, R., Sheppard, K., 2001. Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Available from: SSRN, http://ssrn.com/abstract=1296441.

Engle, R., Shephard, N., Sheppard, K., 2009. Fitting vast dimensional time-varying covariance models. Available from: SSRN, http://ssrn.com/abstract=1354497.

Hafner, C., Franses, P., 2009. A generalized dynamic conditional correlation model: Simulation and application to many assets. Working Paper, Universite Catholique de Louvain, Leuven Belgium.

Hansen, P., Lunde, A., Voev, V., 2010. Realized beta GARCH: A multivariate GARCH model with realized measures of volatility and covolatility. working paper, CREATES, Aarhus University, Denmark.
Jin, X., Maheu, J.M., 2010. Modelling realized covariances and returns. Working Paper, Department of Economics, University of Toronto, Toronto, Ontario, Canada.
Jorion, P., 2006. Value at Risk, third ed. McGraw-Hill.
Noureldin, D., Shephard, N., Sheppard, K., 2011. Multivariate high-frequency-based volatility (HEAVY) models. J. Appl. Econom. forthcoming.
Patton, A., Sheppard, K., 2009. Evaluating volatility and correlation forecasts. In: Andersen, T.G., Davis, R.A., Kreiss, J.P., Mikosch, T. (Eds.), Handbook of Financial Time Series. Springer Verlag, pp. 801–838.
Scholes, M., Williams, J., 1977. Estimating betas from nonsynchronous data. J. Financ. Econ. 5, 309–327.
Tse, Y., Tsui, A., 2002. Multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. J. Bus. Econ. Stat. 20, 351–363.
Voev, V., 2008. Dynamic modelling of large dimensional covariance matrices. In: Bauwens, L., Pohlmeier, W., Veredas, D. (Eds.), High Frequency Financial Econometrics. Physica-Verlag HD, pp. 293–312.

Empirical Exercises

Open the Chapter7Data.xlsx file from the companion site.

1. Calculate daily log returns and plot them on the same scale. How different is the magnitude of variations across the two assets?
2. Compute the unconditional covariance and the correlation for the two assets.
3. Calculate the unconditional 1-day, 1% Value-at-Risk for a portfolio consisting of 50% in each asset. Calculate also the 1-day, 1% Value-at-Risk for each asset individually. Use the normal distribution. Compare the portfolio *VaR* with the sum of individual *VaR*s. What do you see?
4. Estimate an NGARCH(1,1) model for the two assets. Standardize each return using its GARCH standard deviation.
5. Use QMLE to estimate λ in the exponential smoother version of the dynamic conditional correlation (DCC) model for two assets. Set the starting value of λ to 0.94. Calculate the 1-day, 1% *VaR*.
6. Estimate the GARCH DCC model for the two assets. Set the starting values to $\alpha = 0.05$ and $\beta = 0.9$. Plot the dynamic correlations. Calculate and plot the 1-day, 1% *VaR*.

The answers to these exercises can be found in the Chapter7Results.xlsx file, which is available from the companion site.

For more information see the companion site at
http://www.elsevierdirect.com/companions/9780123744487

8 Simulating the Term Structure of Risk

1 Chapter Overview

So far we have focused on the task of computing *VaR* and *ES* for the one-day-ahead horizon only. The dynamic risk models we have introduced have closed-form solutions for one-day-ahead *VaR* and *ES* but not when the horizon of interest is longer than one day. In this case we need to rely on simulation methods for computing *VaR* and *ES*. This chapter introduces two methods for doing so. The simulation-based methods introduced here allow the risk manager to use the dynamic risk model to compute *VaR* and *ES* at any horizon of interest and therefore to compute the entire term structure of risk. By analogy with the term structure of variance plots in Chapter 4 we refer to the term structure of risk as the *VaR* (or *ES*) plotted against the horizon of interest.

The chapter proceeds as follows:

- First, we will consider simulating forward the univariate risk models from Part II of the book. We will introduce two techniques: Monte Carlo simulation, which relies on artificial random numbers, and Filtered Historical Simulation (FHS), which uses historical random shocks.
- Second, we simulate forward in time multivariate risk models with constant correlations across assets. Again we will consider Monte Carlo as well as FHS.
- Third, we simulate multivariate risk models with dynamic correlations using the DCC model from Chapter 7.

We are assuming that the portfolio variance (in the case of univariate risk models) and individual asset variances (in the case of multivariate risk models) have already been modeled and estimated on historical returns using the techniques in Chapters 4 and 5. We are also assuming that the correlation dynamics have been modeled and estimated using the DCC model in Chapter 7.

For convenience we are assuming normally distributed variables when doing Monte Carlo simulation in this chapter. Chapter 9 will provide the details on simulating random variables from the *t* distribution.

Elements of Financial Risk Management. DOI: 10.1016/B978-0-12-374448-7.00008-7

2 The Risk Term Structure in Univariate Models

In the simplistic case, where portfolio returns are normally distributed with a constant variance, σ_{PF}^2, the returns over the next K days are also normally distributed, but with variance $K\sigma_{PF}^2$. In that case, we can easily calculate the *VaR* for returns over the next K days calculated on day t, as

$$VaR_{t+1:t+K}^p = -\sqrt{K}\sigma_{PF}\Phi_p^{-1} = \sqrt{K}VaR_{t+1}^p$$

and similarly *ES* can be computed as

$$ES_{t+1:t+K}^p = \sqrt{K}\sigma_{PF}\frac{\phi\left(\Phi_p^{-1}\right)}{p} = \sqrt{K}ES_{t+1}^p$$

In the much more realistic case where the portfolio variance is time varying, going from 1-day-ahead to K-days-ahead *VaR* is not so simple. As we saw in Chapter 4, the variance of the K-day return is in general

$$\sigma_{t+1:t+K}^2 \equiv E_t\left(\sum_{k=1}^{K} R_{t+k}\right)^2 = \sum_{k=1}^{K} E_t\left[\sigma_{t+k}^2\right]$$

where we have omitted the portfolio, PF, subscripts.

In the simple RiskMetrics variance model, where $\sigma_{t+1}^2 = \lambda\sigma_t^2 + (1-\lambda)R_t^2$, we get

$$\sigma_{t+1:t+K}^2 = \sum_{k=1}^{K}\sigma_{t+1}^2 = K\sigma_{t+1}^2$$

so that variances actually do scale by K in the RiskMetrics model. However, we argued in Chapter 4 that the absence of mean-reversion in variance will imply counterfactual variance forecasts at longer horizons. Furthermore, although the variance is scaled by K in this model, the returns at horizon K are no longer normally distributed. In fact, we can show that the RiskMetrics model implies that returns get further away from normality as the horizon increases, which is counterfactual as we discussed in Chapter 1.

In the symmetric GARCH(1,1) model, where $\sigma_{t+1}^2 = \omega + \alpha R_t^2 + \beta\sigma_t^2$, we instead get

$$\sigma_{t+1:t+K}^2 = K\sigma^2 + \sum_{k=1}^{K}(\alpha+\beta)^{k-1}\left(\sigma_{t+1}^2 - \sigma^2\right) \neq K\sigma_{t+1}^2$$

where

$$\sigma^2 = \frac{\omega}{1-\alpha-\beta}$$

is the unconditional, or average, long-run variance. Recall that in GARCH models tomorrow's variance, σ^2_{t+1}, can conveniently be calculated at the end of today when R_t is realized.

In the GARCH case, the variance does mean revert and it therefore does not scale by the horizon K, and again the returns over the next K days are not normally distributed, even if the 1-day returns are assumed to be. However, a nice feature of mean-reverting GARCH models is that as K gets large, the return distribution does approach the normal. This appears to be a common feature of real-life return data as we argued in Chapter 1.

The upshot is that we are faced with the challenge of computing risk measures such as *VaR* at multiday horizons, without knowing the analytical form for the distribution of returns at those horizons. Fortunately, this challenge can be met through the use of Monte Carlo simulation techniques.

In Chapter 1 we discussed two stylized facts regarding the mean or average daily return—first, that it is very difficult to forecast, and, second that it is very small relative to the daily standard deviation. At a longer horizon, it is still fairly difficult to forecast the mean but its relative importance increases with horizon. Consider a simple example where daily returns are normally distributed with a constant mean and variance as in

$$R_{t+1} \overset{i.i.d.}{\sim} N\left(\mu, \sigma^2\right)$$

The 1-day *VaR* is thus

$$VaR^p_{t+1} = -\left(\mu + \sigma\Phi^{-1}_p\right) \approx -\sigma\Phi^{-1}_p$$

where the last equation holds approximately because the daily mean is typically orders of magnitude smaller than the standard deviation as we saw in Chapter 1.

The K-day return in this case is distributed as

$$R_{t+1:t+K} \sim N\left(K\mu, K\sigma^2\right)$$

and the K-day *VaR* is thus

$$VaR^p_{t+1:t+K} = -\left(K\mu + \sqrt{K}\sigma\Phi^{-1}_p\right) \not\approx -\sqrt{K}\sigma\Phi^{-1}_p$$

As the horizon, K, gets large, the relative importance of the mean increases and the zero-mean approximation no longer holds. Similarly, for *ES*

$$ES^p_{t+1:t+K} = -\left(K\mu + \sqrt{K}\sigma\frac{\phi\left(\Phi^{-1}_p\right)}{p}\right) \not\approx \sqrt{K}\sigma\frac{\phi\left(\Phi^{-1}_p\right)}{p}$$

Although the mean return is potentially important at longer horizons, in order to save on notation, we will still assume that the mean is zero in the sections that follow. However, it is easy to generalize the analysis to include a nonzero mean.

2.1 Monte Carlo Simulation

We illustrate the power of Monte Carlo simulation (MCS) through a simple example. Consider our GARCH(1,1)-normal model of returns, where

$$R_{t+1} = \sigma_{t+1} z_{t+1}, \quad \text{with } z_{t+1} \overset{i.i.d.}{\sim} N(0,1)$$

and

$$\sigma_{t+1}^2 = \omega + \alpha R_t^2 + \beta \sigma_t^2$$

As mentioned earlier, at the end of day t we obtain R_t and we can calculate σ_{t+1}^2, which is tomorrow's variance in the GARCH model.

Using random number generators, which are standard in most quantitative software packages, we can generate a set of artificial (or pseudo) random numbers

$$\breve{z}_{i,1}, \quad i = 1, 2, \ldots, MC$$

drawn from the standard normal distribution, $N(0,1)$. MC denotes the number of draws, which should be large, for example, 10,000. To confirm that the random numbers do indeed conform to the standard normal distribution, a QQ plot of the random numbers can be constructed.

From these random numbers we can calculate a set of hypothetical returns for tomorrow as

$$\breve{R}_{i,t+1} = \sigma_{t+1} \breve{z}_{i,1}$$

Given these hypothetical returns, we can update the variance to get a set of hypothetical variances for the day after tomorrow, $t+2$, as follows:

$$\breve{\sigma}_{i,t+2}^2 = \omega + \alpha \breve{R}_{i,t+1}^2 + \beta \sigma_{t+1}^2$$

Given a new set of random numbers drawn from the $N(0,1)$ distribution,

$$\breve{z}_{i,2}, \quad i = 1, 2, \ldots, MC$$

we can calculate the hypothetical return on day $t+2$ as

$$\breve{R}_{i,t+2} = \breve{\sigma}_{i,t+2} \breve{z}_{i,2}$$

and the variance is now updated using

$$\breve{\sigma}_{i,t+3}^2 = \omega + \alpha \breve{R}_{i,t+2}^2 + \beta \breve{\sigma}_{i,t+2}^2$$

Graphically, we can illustrate the simulation of hypothetical daily returns from day $t+1$ to day $t+K$ as

$$\begin{array}{cccccc} & \breve{z}_{1,1} \to \breve{R}_{1,t+1} \to \breve{\sigma}^2_{1,t+2} & \breve{z}_{1,2} \to \breve{R}_{1,t+2} \to \breve{\sigma}^2_{1,t+3} & \cdots & \breve{z}_{1,K} \to \breve{R}_{1,t+K} \\ \nearrow & \breve{z}_{2,1} \to \breve{R}_{2,t+1} \to \breve{\sigma}^2_{2,t+2} & \breve{z}_{2,2} \to \breve{R}_{2,t+2} \to \breve{\sigma}^2_{2,t+3} & \cdots & \breve{z}_{2,K} \to \breve{R}_{2,t+K} \\ \sigma^2_{t+1} \longrightarrow & \cdots & \cdots & \cdots & \cdots \\ \searrow & \cdots & \cdots & \cdots & \cdots \\ & \breve{z}_{MC,1} \to \breve{R}_{MC,t+1} \to \breve{\sigma}^2_{MC,t+2} & \breve{z}_{MC,2} \to \breve{R}_{MC,t+2} \to \breve{\sigma}^2_{MC,t+3} & \cdots & \breve{z}_{MC,K} \to \breve{R}_{MC,t+K} \end{array}$$

Each row corresponds to a so-called Monte Carlo simulation path, which branches out from σ^2_{t+1} on the first day, but which does not branch out after that. On each day a given branch gets updated with a new random number, which is different from the one used any of the days before. We end up with MC sequences of hypothetical daily returns for day $t+1$ through day $t+K$. From these hypothetical future daily returns, we can easily calculate the hypothetical K-day return from each Monte Carlo path as

$$\breve{R}_{i,t+1:t+K} = \sum_{k=1}^{K} \breve{R}_{i,t+k}, \quad \text{for } i = 1, 2, \ldots, MC$$

If we collect these MC hypothetical K-day returns in a set $\left\{\breve{R}_{i,t+1:t+K}\right\}_{i=1}^{MC}$, then we can calculate the K-day value at risk simply by calculating the $100p$th percentile as in

$$VaR^p_{t+1:t+K} = -Percentile\left\{\left\{\breve{R}_{i,t+1:t+K}\right\}_{i=1}^{MC}, 100p\right\}$$

We can also use Monte Carlo to compute the expected shortfall at different horizons

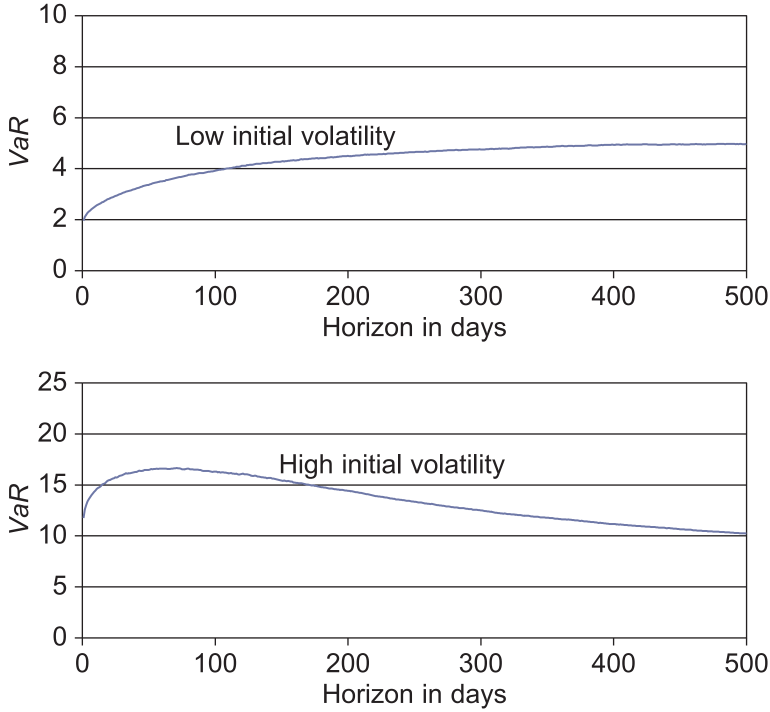
$$ES^p_{t+1:t+K} = -\frac{1}{p \cdot MC} \sum_{i=1}^{MC} \breve{R}_{i,t+1:t+K} \cdot \mathbf{1}\left(\breve{R}_{i,t+1:t+K} < -VaR^p_{t+1:t+K}\right)$$

where $\mathbf{1}(\bullet)$ takes the value 1 if the argument is true and zero otherwise.

Notice that in contrast to the HS and WHS techniques introduced in Chapter 2, the GARCH-MCS method outlined here is truly conditional in nature as it builds on today's estimate of tomorrow's variance, σ^2_{t+1}.

A key advantage of the MCS technique is its flexibility. We can use MCS for any assumed distribution of standardized returns—normality is not required. If we think the standardized $t(d)$ distribution with $d = 12$ for example describes the data better, then we simply draw from this distribution. Commercial software packages typically contain the regular $t(d)$ distribution, but we can standardize these draws by multiplying by $\sqrt{(d-2)/d}$ as we saw in Chapter 6. Furthermore, the MCS technique can be used for any fully specified dynamic variance model.

Figure 8.1 *VaR* term structures using NGARCH and Monte Carlo simulation.

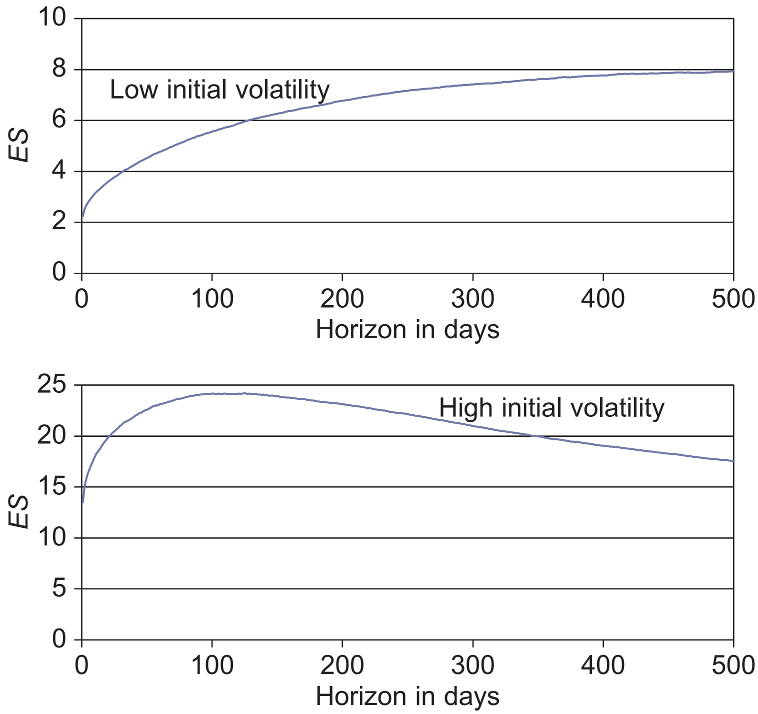


Notes: The top panel shows the S&P 500 *VaR* per day across horizons when the current volatility is one-half its long-run value. The bottom panel assumes the current volatility is three times its long-run value. The *VaR* is simulated using Monte Carlo on an NGARCH model.

In Figure 8.1 we apply the NGARCH model for the S&P 500 from Chapter 4 along with a normal distribution assumption. We use Monte Carlo simulation to construct and plot *VaR* per day, $VaR^p_{t+1:t+K}/\sqrt{K}$, as a function of horizon K for two different values of σ_{t+1}. In the top panel the initial volatility is one-half the unconditional level and in the bottom panel σ_{t+1} is three times the unconditional level. The horizon goes from 1 to 500 trading days, corresponding roughly to two calendar years.

The *VaR* coverage level p is set to 1%. Figure 8.1 gives a *VaR*-based picture of the term structure of risk. Perhaps surprisingly the term structure of *VaR* is initially upward sloping both when volatility is low and when it is high. The *VaR* term structure is driven partly by the variance term structure, which is upward sloping when current volatility is low and downward sloping when current volatility is high as we saw in Chapter 4. But the *VaR* term structure is also driven by the term structure of skewness and kurtosis and other moments. Kurtosis is strongly increasing at short horizons and then decreasing for longer horizons. This hump-shape in the term structure of kurtosis creates the hump in the *VaR* that we see in the bottom panel of Figure 8.1 when the initial volatility is high.

Figure 8.2 *ES* term structures using NGARCH and Monte Carlo simulation.



Notes: The top panel shows the S&P 500 *ES* per day across horizons when the current volatility is one-half its long-run value. The bottom panel assumes the current volatility is three times its long-run value. The *ES* is simulated using Monte Carlo on an NGARCH model.

In Figure 8.2 we plot the $ES^p_{t+1:t+K}$ per day, $ES^p_{t+1:t+K}/\sqrt{K}$, against horizon K. The coverage level p is again set to 1% and the horizon goes from 1 to 500 trading days. Figure 8.2 gives an *ES*-based picture of the term structure of risk, which is clearly qualitatively similar to the term structure of *VaR* in Figure 8.1. Note however, that the slope of the *ES* term structure in the upper panel of Figure 8.2 is steeper than the corresponding *VaR* term structure in the upper panel of Figure 8.2. Note also that the hump in the *ES* term structure in the bottom panel of Figure 8.2 is more pronounced than the hump in the *VaR* term structure in the upper panel of Figure 8.1.

2.2 Filtered Historical Simulation (FHS)

In the book so far, we have discussed methods that take very different approaches: Historical Simulation (HS) in Chapter 2 is a completely model-free approach, which imposes virtually no structure on the distribution of returns: the historical returns calculated with today's weights are used directly to calculate a percentile. The GARCH Monte Carlo simulation (MCS) approach in this chapter takes the opposite view and

assumes parametric models for variance, correlation (if a disaggregate model is estimated), and the distribution of standardized returns. Random numbers are then drawn from this distribution to calculate the desired risk measure.

Both of these extremes in the model-free/model-based spectrum have pros and cons. Taking a model-based approach (MCS, for example) is good if the model is a fairly accurate description of reality. Taking a model-free approach (HS, for example) is sensible in that the observed data may capture features of the returns distribution that are not captured by any standard parametric model.

The Filtered Historical Simulation (FHS) approach, which we introduced in Chapter 6, attempts to combine the best of the model-based with the best of the model-free approaches in a very intuitive fashion. FHS combines model-based methods of variance with model-free methods of the distribution of shocks.

Assume we have estimated a GARCH-type model of our portfolio variance. Although we are comfortable with our variance model, we are not comfortable making a specific distributional assumption about the standardized returns, such as a normal or a $\tilde{t}(d)$ distribution. Instead, we would like the past returns data to tell us about the distribution directly without making further assumptions.

To fix ideas, consider again the simple example of a GARCH(1,1) model:

$$R_{t+1} = \sigma_{t+1} z_{t+1}$$

where

$$\sigma_{t+1}^2 = \omega + \alpha R_t^2 + \beta \sigma_t^2$$

Given a sequence of past returns, $\{R_{t+1-\tau}\}_{\tau=1}^m$, we can estimate the GARCH model and calculate past standardized returns from the observed returns and from the estimated standard deviations as

$$\hat{z}_{t+1-\tau} = R_{t+1-\tau}/\sigma_{t+1-\tau}, \quad \text{for } \tau = 1, 2, \ldots, m$$

We will refer to the set of standardized returns as $\left\{\hat{z}_{t+1-\tau}\right\}_{\tau=1}^m$. The number of historical observations, m, should be as large as possible.

Moving forward now, at the end of day t we obtain R_t and we can calculate σ_{t+1}^2, which is day $t+1$'s variance in the GARCH model. Instead of drawing random $\hat{z}$s from a random number generator, which relies on a specific distribution, we can draw with replacement from our own database of past standardized residuals, $\left\{\hat{z}_{t+1-\tau}\right\}_{\tau=1}^m$. The random drawing can be operationalized by generating a discrete uniform random variable distributed from 1 to m. Each draw from the discrete distribution then tells us which τ and thus which $\hat{z}_{t+1-\tau}$ to pick from the set $\left\{\hat{z}_{t+1-\tau}\right\}_{\tau=1}^m$.

We again build up a distribution of hypothetical future returns as

$$\begin{array}{llllll}
 & \hat{z}_{1,1} \to \hat{R}_{1,t+1} \to \hat{\sigma}^2_{1,t+2} & \hat{z}_{1,2} \to \hat{R}_{1,t+2} \to \hat{\sigma}^2_{1,t+3} & \ldots & \hat{z}_{1,K} \to \hat{R}_{1,t+K} \\
\nearrow & \hat{z}_{2,1} \to \hat{R}_{2,t+1} \to \hat{\sigma}^2_{2,t+2} & \hat{z}_{2,2} \to \hat{R}_{2,t+2} \to \hat{\sigma}^2_{2,t+3} & \ldots & \hat{z}_{2,K} \to \hat{R}_{2,t+K} \\
\sigma^2_{t+1} \longrightarrow & \ldots & \ldots & \ldots & \ldots \\
\searrow & \ldots & \ldots & \ldots & \ldots \\
 & \hat{z}_{FH,1} \to \hat{R}_{FH,t+1} \to \hat{\sigma}^2_{FH,t+2} & \hat{z}_{FH,2} \to \hat{R}_{FH,t+2} \to \hat{\sigma}^2_{FH,t+3} & \ldots & \hat{z}_{FH,K} \to \hat{R}_{FH,t+K}
\end{array}$$

where FH is the number of times we draw from the standardized residuals on each future date, for example 10,000, and where K is the horizon of interest measured in number of days.

We end up with FH sequences of hypothetical daily returns for day $t+1$ through day $t+K$. From these hypothetical daily returns, we calculate the hypothetical K-day returns as

$$\hat{R}_{i,t+1:t+K} = \sum_{k=1}^{K} \hat{R}_{i,t+k}, \qquad \text{for } i = 1, 2, \ldots, FH$$

If we collect the FH hypothetical K-day returns in a set $\left\{\hat{R}_{i,t+1:t+K}\right\}_{i=1}^{FH}$, then we can calculate the K-day Value-at-Risk simply by calculating the 100pth percentile as in

$$VaR^p_{t+1:t+K} = -Percentile\left\{\left\{\hat{R}_{i,t+1:t+K}\right\}_{i=1}^{FH}, 100p\right\}$$

The ES measure can again be calculated from the simulated returns by simply taking the average of all the $\hat{R}_{i,t+1:t+K}$s that fall below the $-VaR^p_{t+1:t+K}$ number; that is,

$$ES^p_{t+1:t+K} = -\frac{1}{p \cdot FH} \cdot \sum_{i=1}^{FH} \hat{R}_{i,t+1:t+K} \cdot \mathbf{1}\left(\hat{R}_{i,t+1:t+K} < -VaR^p_{t+1:t+K}\right)$$

where as before the indicator function $\mathbf{1}(\bullet)$ returns a 1 if the argument is true and zero if not.

An interesting and useful feature of FHS as compared with simple HS is that it can generate large losses in the forecast period, even without having observed a large loss in the recorded past returns. Consider the case where we have a relatively large negative z in our database, which occurred on a relatively low variance day. If this z gets combined with a high variance day in the simulation period then the resulting hypothetical loss will be large.

In Figure 8.3 we use the FHS approach based on the NGARCH model for the S&P 500 returns. We use the NGARCH-FHS model to construct and plot the $VaR^p_{t+1:t+K}$ per day as a function of horizon K for two different values of σ_{t+1}. In the top panel the initial volatility is one-half the unconditional level and in the bottom panel σ_{t+1} is three times the unconditional level. The horizons goes from 1 to 500 trading days, corresponding roughly to two calendar years.

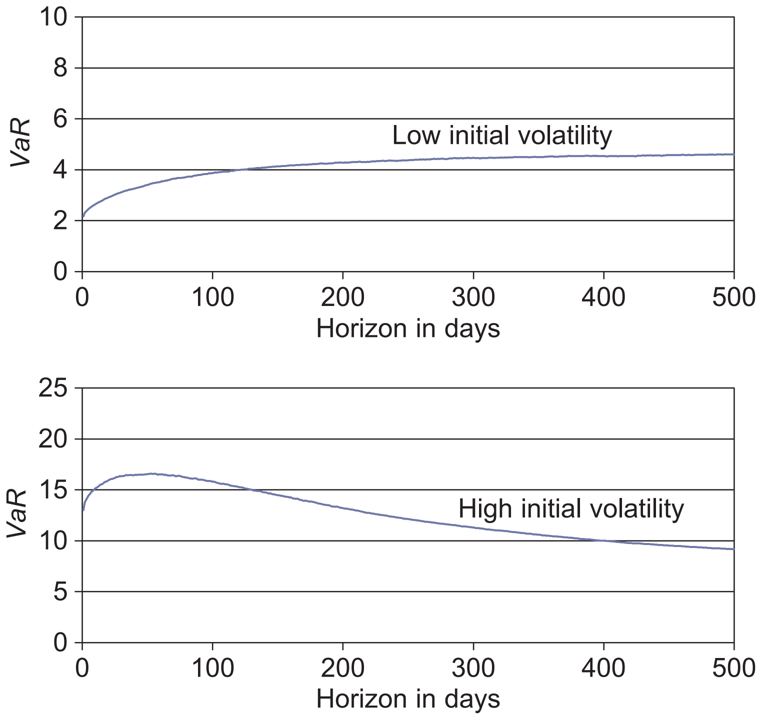
The VaR coverage level p is set to 1% again. Comparing Figure 8.3 with Figure 8.1 we see that for this S&P 500 portfolio the Monte Carlo and FHS simulation methods give roughly equal VaR term structures when the initial volatility is the same.

In Figure 8.4 we plot the $ES^p_{t+1:t+K}$ per day against horizon K.

The coverage level p is again set to 1% and the horizon goes from 1 to 500 trading days. The FHS-based ES term structure in Figure 8.4 closely resembles the NGARCH Monte Carlo-based ES term structure in Figure 8.2.

We close this section by reemphasizing that the FHS method suggested here combines a conditional model for variance with a Historical Simulation method for the

Figure 8.3 *VaR* term structures using NGARCH and filtered Historical Simulation.



Notes: The top panel shows the S&P 500 *VaR* per day across horizons when the current volatility is one-half its long-run value. The bottom panel assumes the current volatility is three times its long run value. The *VaR* is simulated using FHS on an NGARCH model.

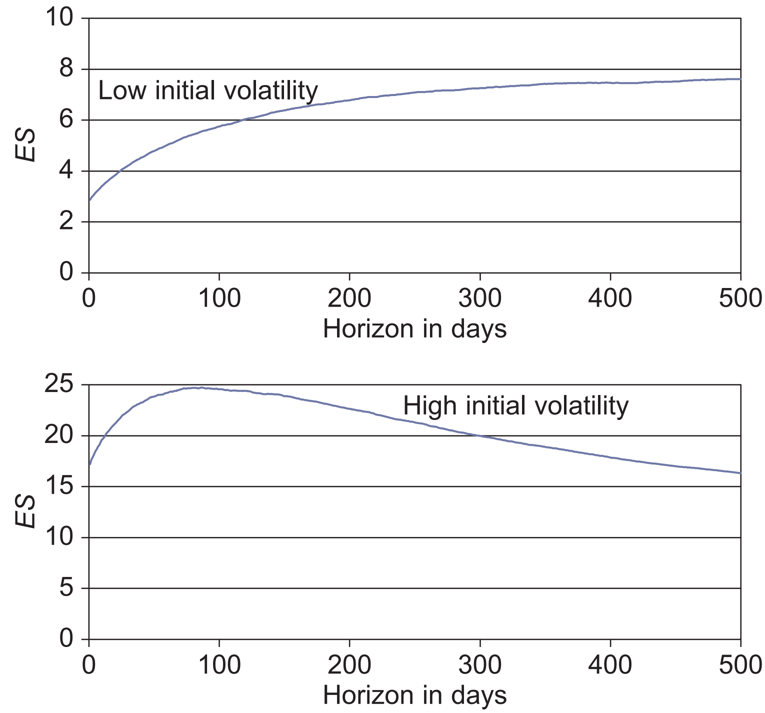
standardized returns. FHS thus captures the current level of market volatility via σ_{t+1} but we do not need to make assumptions about the tail distribution. The FHS method has been found to perform very well in several studies and it should be given serious consideration by any risk management team.

3 The Risk Term Structure with Constant Correlations

The univariate methods discussed in Section 2 are useful if the main purpose of the risk model is risk measurement. If instead the model is required for active risk management including deciding on optimal portfolio allocations, or *VaR* sensitivities to allocation changes, then a multivariate model is required. In this section, we use the multivariate models built in Chapter 7 to simulate *VaR* and *ES* for different maturities. The multivariate risk models allow us to compute risk measures for different hypothetical portfolio allocations without having to reestimate model parameters.

We will assume that the risk manager knows his or her set of assets of interest. This set can either contain all the assets in the portfolio or a smaller set of base assets, which

Figure 8.4 *ES* term structures using NGARCH and filtered Historical Simulation.



Notes: The top panel shows the S&P 500 *ES* per day across horizons when the current volatility is one-half its long-run value. The bottom panel assumes the current volatility is three times its long-run value. The *ES* is simulated using FHS on an NGARCH model.

are believed to be the main drivers of risk in the portfolio. Base asset choices are, of course, portfolio-specific, but typical examples include equity indices, bond indices, and exchange rates as well as more fundamental economic drivers such as oil prices and real estate prices as discussed in Chapter 7.

Once the set of assets has been determined, the next step in the multivariate model is to estimate a dynamic volatility model of the type in Chapters 4 and 5 for each of the n assets. When this is complete, we can write the n asset returns in vector form as follows:

$$r_{t+1} = D_{t+1} z_{t+1}$$

where D_{t+1} is an $n \times n$ diagonal matrix containing the dynamic standard deviations on the diagonal, and zeros on the off diagonal. The $n \times 1$ vector z_{t+1} contains the shocks from the dynamic volatility model for each asset.

Now, define the conditional covariance matrix of the returns as

$$Var_t(r_{t+1}) = \Sigma_{t+1} = D_{t+1} \Upsilon D_{t+1}$$

where Υ is a constant $n \times n$ matrix containing the base asset correlations on the off diagonals and ones on the diagonal. Later we will consider DCC models where the correlation matrix is time varying.

When simulating the multivariate model forward we face a new challenge, namely, that we must ensure that the vector of shocks have the correct correlation matrix, Υ. Random number generators provide us with uncorrelated random standard normal variables, z_t^u, and we must correlate them before using them to simulate returns forward.

In the case of two uncorrelated shocks, we have

$$E\left[z_{t+1}^u \left(z_{t+1}^u\right)'\right] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

but we want to create correlated shocks with the correlation matrix

$$E\left[z_{t+1} \left(z_{t+1}\right)'\right] = \Upsilon = \begin{bmatrix} 1 & \rho_{1,2} \\ \rho_{1,2} & 1 \end{bmatrix}$$

We therefore need to find the matrix square root, $\Upsilon^{1/2}$, so that $\Upsilon^{1/2}\left(\Upsilon^{1/2}\right)' = \Upsilon$ and so that $z_{t+1} = \Upsilon^{1/2} z_{t+1}^u$ will give the correct correlation matrix, namely

$$E\left[z_{t+1} \left(z_{t+1}\right)'\right] = E\left[\Upsilon^{1/2} z_{t+1}^u \left(z_{t+1}^u\right)' \left(\Upsilon^{1/2}\right)'\right] = \Upsilon$$

In the bivariate case we have that

$$\Upsilon^{1/2} = \begin{bmatrix} 1 & 0 \\ \rho_{1,2} & \sqrt{1-\rho_{1,2}^2} \end{bmatrix}$$

so that when multiplying out $z_{t+1} = \Upsilon^{1/2} z_{t+1}^u$ we get

$$z_{1,t+1} = z_{1,t+1}^u$$
$$z_{2,t+1} = \rho_{1,2} z_{1,t+1}^u + \sqrt{1-\rho_{1,2}^2}\, z_{2,t+1}^u$$

which implies that

$$E\left[z_{1,t+1}\right] = E\left[z_{1,t+1}^u\right] = 0$$
$$E\left[z_{2,t+1}\right] = \rho_{1,2} E\left[z_{1,t+1}^u\right] + \sqrt{1-\rho_{1,2}^2}\, E\left[z_{2,t+1}^u\right] = 0$$

and

$$Var\left[z_{1,t+1}\right] = Var\left[z_{1,t+1}^u\right] = 1$$
$$Var\left[z_{2,t+1}\right] = \rho_{1,2}^2 Var\left[z_{1,t+1}^u\right] + \left(1-\rho_{1,2}^2\right) Var\left[z_{2,t+1}^u\right] = 1$$

because $Var\left[z^u_{1,t+1}\right] = Var\left[z^u_{2,t+1}\right] = 1$. Thus $z_{1,t+1}$ and $z_{2,t+1}$ will each have a mean of 0 and a variance of 1 as desired. Finally we can check the correlation. We have

$$E\left[z_{1,t+1}z_{2,t+1}\right] = \rho_{1,2}E\left[z^u_{1,t+1}z^u_{1,t+1}\right] + \sqrt{1-\rho^2_{1,2}}E\left[z^u_{1,t+1}z^u_{2,t+1}\right] = \rho_{1,2}$$

so that the shocks will have a correlation of $\rho_{1,2}$ as desired.

We can also verify the $\Upsilon^{1/2}$ matrix by multiplying it by its transpose

$$\Upsilon^{1/2}\left(\Upsilon^{1/2}\right)' = \begin{bmatrix} 1 & 0 \\ \rho_{1,2} & \sqrt{1-\rho^2_{1,2}} \end{bmatrix} \begin{bmatrix} 1 & \rho_{1,2} \\ 0 & \sqrt{1-\rho^2_{1,2}} \end{bmatrix} = \begin{bmatrix} 1 & \rho_{1,2} \\ \rho_{1,2} & 1 \end{bmatrix} = \Upsilon$$

In the case of $n > 2$ assets we need to use a so-called Cholesky decomposition or a spectral decomposition of Υ to compute $\Upsilon^{1/2}$. See the references for details on these methods.

3.1 Multivariate Monte Carlo Simulation

In order to simulate the model forward in time using Monte Carlo we need to assume a multivariate distribution of the vector of shocks, $\check{z}$. In this chapter we will rely on the multivariate standard normal distribution because it is convenient and so allows us to focus on the issues involved in simulation. In Chapter 9 we will look at more complicated multivariate t distributions.

The algorithm for multivariate Monte Carlo simulation is as follows:

- First, draw a vector of uncorrelated random normal variables $\check{z}^u_{i,1}$ with a mean of zero and a variance of one.
- Second, use the matrix square root $\Upsilon^{1/2}$ to correlate the random variables; this gives $\check{z}_{i,t+1} = \Upsilon^{1/2}\check{z}^u_{i,1}$.
- Third, update the variances for each asset using the approach in Section 2.
- Fourth, compute returns for each asset using the approach in Section 2.

Loop through these four steps from day $t+1$ until day $t+K$. Now we can compute the portfolio return using the known portfolio weights and the vector of simulated returns on each day.

Repeating these steps $i = 1, 2, \ldots, MC$ times gives a Monte Carlo distribution of portfolio returns. From these MC portfolio returns we can compute VaR and ES from the simulated portfolio returns as in Section 2.

3.2 Multivariate Filtered Historical Simulation

Multivariate Filtered Historical Simulation can be done easily when we assume constant correlations.

- First, draw a vector (across assets) of historical shocks from a particular day in the historical sample of shocks, and use that to simulate tomorrow's shock, $\hat{z}_{i,1}$. The key insight is that when we draw the entire vector (across assets) of historical shocks from the same day, they will preserve the correlation across assets that existed historically as long as correlations are constant over time.
- Second, update the variances for each asset using the approach in Section 2.
- Third, compute returns for each asset using the approach in Section 2.

Loop through these steps from day $t+1$ until day $t+K$. Now we can compute the portfolio return using the known portfolio weights and the vector of simulated returns on each day as before.

Repeating these steps $i = 1, 2, \ldots, FH$ times gives a simulated distribution of portfolio returns. From these FH portfolio returns we can compute VaR and ES from the simulated portfolio returns as in Section 2.

4 The Risk Term Structure with Dynamic Correlations

We now consider the more complicated case where the correlations are dynamic as in the DCC model in Chapter 7. We again have

$$r_{t+1} = D_{t+1} z_{t+1}$$

where D_{t+1} is an $n \times n$ diagonal matrix containing the GARCH standard deviations on the diagonal, and zeros on the off diagonal. The $n \times 1$ vector z_t contains the shocks from the GARCH models for each asset.

Now, we have

$$Var_t(r_{t+1}) = \Sigma_{t+1} = D_{t+1} \Upsilon_{t+1} D_{t+1}$$

where Υ_{t+1} is an $n \times n$ matrix containing the base asset correlations on the off diagonals and ones on the diagonal. The elements in D_{t+1} can be simulated forward using the methods in Section 2 but we now also need to simulate the correlation matrix forward.

4.1 Monte Carlo Simulation with Dynamic Correlations

As mentioned before, random number generators typically provide us with uncorrelated random standard normal variables, $\check{z}^u$, and we must correlate them before simulating returns forward.

At the end of day t the GARCH and DCC models provide us with D_{t+1} and Υ_{t+1} without having to do simulation. We can therefore compute a random return for day $t+1$ as

$$\check{r}_{i,t+1} = D_{t+1} \Upsilon_{t+1}^{1/2} \check{z}_{i,1}^{u} = D_{t+1} \check{z}_{i,t+1}$$

where $\check{z}_{i,t+1} = \Upsilon_{t+1}^{1/2} \check{z}_{i,1}^{u}$.

Using the new simulated shock vector, $\check{z}_{i,t+1}$, we can update the volatilities and correlations using the GARCH models and the DCC model. We thus obtain simulated $\check{D}_{i,t+2}$ and $\check{\Upsilon}_{i,t+2}$. Drawing a new vector of uncorrelated shocks, $\check{z}^u_{i,2}$, enables us to simulate the return for the second day ahead as

$$\check{r}_{i,t+2} = \check{D}_{i,t+2}\check{\Upsilon}^{1/2}_{i,t+2}\check{z}^u_{i,2} = \check{D}_{i,t+2}\check{z}_{i,t+2}$$

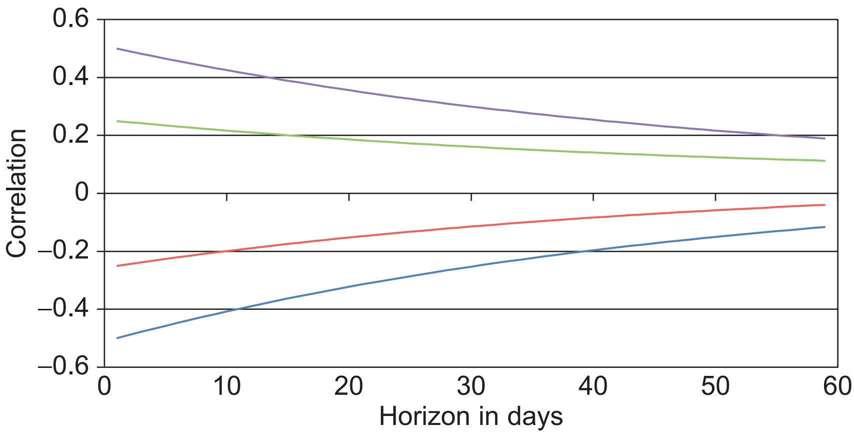
where $\check{z}_{i,t+2} = \check{\Upsilon}^{1/2}_{i,t+2}\check{z}^u_{i,2}$.

We continue this simulation from day $t+1$ through day $t+K$, and repeat it for $i = 1, 2, \ldots, MC$ vectors of simulated shocks on each day. As before we can compute the portfolio return using the known portfolio weights and the vector of simulated returns on each day. From these MC portfolio returns we can compute VaR and ES from the simulated portfolio returns as in Section 2.

In Figure 8.5 we use the DCC model for S&P 500 returns and the 10-year treasury bond index from Chapter 7 to plot the expected future correlations. We have assumed four different values of the current correlation, ranging from −0.5 in the blue line to +0.5 in the purple line. Note that over the 60-day horizon considered, the correlations converge toward the long-run correlation value but significant differences remain even after 60 days.

In Chapter 4 we saw how the expected future variance can be computed analytically from current variance using the GARCH model dynamics. The key equations were repeated in Section 2. Unfortunately for dynamic correlation models, such exact analytical formulas for expected future correlation do not exist. We need to rely on the simulation methods developed here in order to construct correlation forecasts for more than one day ahead. If we for example want to construct a forecast for the correlation

Figure 8.5 DCC correlation forecasts by Monte Carlo simulation.



Notes: The correlation forecast across horizons is shown for four different levels of current correlation. The forecasts are computed using Monte Carlo simulation on the DCC model for the S&P 500 and 10-year treasury bond.

matrix two days ahead we can use

$$\Upsilon_{t+2|t} = \frac{1}{MC}\sum_{i=1}^{MC} \check{\Upsilon}_{i,t+2}$$

where the Monte Carlo average is done element by element for each of the correlations in the matrix.

4.2 Filtered Historical Simulation with Dynamic Correlations

When correlations across assets are assumed to be constant then FHS is relatively easy because we can draw from historical asset shocks, using the entire vector (across assets) of historical shocks. The (constant) historical correlation will be preserved in the simulated shocks. When correlations are dynamic then we need to ensure that the correlation dynamics are simulated forward but in FHS we still want to use the historical shocks.

In this case we must first create a database of historical dynamically uncorrelated shocks from which we can resample. We create the dynamically uncorrelated historical shock as

$$\hat{z}^u_{t+1-\tau} = \Upsilon^{-1/2}_{t+1-\tau}\hat{z}_{t+1-\tau}, \quad \text{for } \tau = 1, 2, \ldots, m$$

where $\hat{z}_{t+1-\tau}$ is the vector of standardized shocks on day $t+1-\tau$ and where $\Upsilon^{-1/2}_{t+1-\tau}$ is the inverse of the matrix square-root of the conditional correlation matrix $\Upsilon_{t+1-\tau}$.

When calculating the multiday conditional *VaR* and *ES* from the model, we again need to simulate daily returns forward from today's (day t) forecast of tomorrow's matrix of volatilities, D_{t+1} and correlations, Υ_{t+1}.

From the database of uncorrelated shocks $\left\{\hat{z}^u_{t+1-\tau}\right\}^m_{\tau=1}$, we can draw a random vector of historical uncorrelated shocks, called $\hat{z}^u_{i,1}$. It is important to note that in order to preserve asset-specific characteristics and potential nonlinear dependence in the shocks, we draw an entire vector representing the same day for all the assets.

From this draw, we can compute a random return for day $t+1$ as

$$\hat{r}_{i,t+1} = D_{t+1}\Upsilon^{1/2}_{t+1}\hat{z}^u_{i,1} = D_{t+1}\hat{z}_{i,t+1}$$

where $\hat{z}_{i,t+1} = \Upsilon^{1/2}_{t+1}\hat{z}^u_{i,1}$.

Using the new simulated shock vector, $\hat{z}_{i,t+1}$, we can update the volatilities and correlations using the GARCH models and the DCC model. We thus obtain simulated $\hat{D}_{i,t+2}$ and $\hat{\Upsilon}_{i,t+2}$. Drawing a new vector of uncorrelated shocks, $\hat{z}^u_{i,2}$, enables us to simulate the return for the second day as

$$\hat{r}_{i,t+2} = \hat{D}_{i,t+2}\hat{\Upsilon}^{1/2}_{i,t+2}\hat{z}^u_{i,2} = D_{t+2}\hat{z}_{i,t+2}$$

where $\hat{z}_{i,t+2} = \hat{\Upsilon}^{1/2}_{t+1}\hat{z}^u_{i,2}$. We continue this simulation for K days, and repeat it for *FH* vectors of simulated shocks on each day. As before we can compute the portfolio

return using the known portfolio weights and the vector of simulated returns on each day. From these *FH* portfolio returns we can compute *VaR* and *ES* from the simulated portfolio returns as in Section 2.

The advantages of the multivariate FHS approach tally with those of the univariate case: It captures current market conditions by means of dynamic variance and correlation models. It makes no assumption on the conditional multivariate shock distributions. And, it allows for the computation of any risk measure for any investment horizon of interest.

5 Summary

Risk managers rarely have one particular horizon of interest but rather want to know the risk profile across many different horizons; that is, the term structure of risk. The purpose of this chapter has therefore been to introduce Monte Carlo simulation and filtered Historical Simulation techniques, which can be used to compute the term structure of risk in the univariate risk models in Part II as well as in the multivariate risk models in Chapter 7. It is important to keep in mind that because we are simulating from dynamic risk models, we use all the relevant information available at any given time to compute the risk forecasts across future horizons.

Chapter 7 assumed the multivariate normal distribution. This assumption was made for convenience and not for realism. We need to develop nonnormal multivariate distributions that can be used in risk computation across different horizons as well. This is the task of the upcoming Chapter 9.

Further Resources

Theoretical issues involved in temporal aggregation of GARCH models are analyzed in Drost and Nijman (1993). Diebold et al. (1998a) study the problems arising in risk management from simple scaling rules of variance across horizons. Christoffersen et al. (1998) elaborate on the issues involved in calculating *VaR*s at different horizons. Christoffersen and Diebold (2000) investigate the usefulness of dynamic variance models for risk management at various forecast horizons. Portfolio aggregation of GARCH models is analyzed in Zaffaroni (2007).

A thorough and current treatment of Monte Carlo methods in financial engineering can be found in the Glasserman (2004) book. Hammersley and Handscomb (1964) is the classic reference on Monte Carlo methods.

Diebold et al. (1998a); Hull and White (1998); and Barone-Adesi et al. (1999) independently suggested the filtered Historical Simulation approach. See also Barone-Adesi et al. (1998); and Barone-Adesi et al. (2002), who consider an application of FHS to portfolios of options and futures. Pritsker (2006) provides a powerful comparison between FHS and traditional Historical Simulation.

When constructing correlated random shocks, Patton and Sheppard (2009) recommend the spectral decomposition of the correlation matrix over the standard Cholesky

decomposition because the latter is not invariant to the ordering of the assets in the vector of shocks.

Engle (2009) and Engle and Sheppard (2001) develop approximate formulas for correlation forecasts in DCC models. Asai and McAleer (2009) consider a stochastic correlation modeling approach.

Parametric alternatives to the Filtered Historical Simulation approach include specifying a multivariate normal or *t* distribution for the GARCH shocks. See, for example, Pesaran et al. (2009) as well as Chapter 9 in this book.

See Engle and Manganelli (2004) for a survey of different *VaR* modeling approaches. Manganelli (2004) considers a unique asset allocation approach that only requires a univariate model.

References

Asai, M., McAleer, M., 2009. The structure of dynamic correlations in multivariate stochastic volatility models. J. Econom. 150, 182–192.

Barone-Adesi, G., Bourgoin, F., Giannopoulos, K., 1998. Don't look back. Risk 11, August, 100–104.

Barone-Adesi, G., Giannopoulos, K., Vosper, L., 1999. VaR without correlations for non-linear portfolios. J. Futures Mark. 19, 583–602.

Barone-Adesi, G., Giannopoulos, K., Vosper, L., 2002. Backtesting derivative portfolios with filtered historical simulation (FHS). Eur. Financ. Manag. 8, 31–58.

Christoffersen, P., Diebold, F., 2000. How relevant is volatility forecasting for financial risk management? Rev. Econ. Stat. 82, 1–11.

Christoffersen, P., Diebold, F., Schuermann, T., 1998. Horizon problems and extreme events in financial risk management. Fed. Reserve Bank New York Econ. Policy Rev. 4, 109–118.

Diebold, F.X., Hickman, A., Inoue, A., Schuermann, T., 1998a. Scale models. Risk 11, 104–107.

Drost, F., Nijman, T., 1993. Temporal aggregation of GARCH processes. Econometrica 61, 909–927.

Engle, R., 2009. Anticipating Correlations: A New Paradigm for Risk Management. Princeton University Press, Princeton, NJ.

Engle, R., Manganelli, S., 2004. A comparison of value at risk models in finance. In: Szego, G. (Ed.), Risk Measures for the 21st Century. Wiley Finance, John Wiley Sons, Ltd., Chichester, West Sussex, England, pp. 123–144.

Engle, R., Sheppard, K., 2001. Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Available from: SSRN, http://ssrn.com/abstract=1296441.

Glasserman, P., 2004. Monte Carlo Methods in Financial Engineering. Springer Verlag.

Hammersley, J., Handscomb, D., 1964. Monte Carlo Methods. Fletcher and Sons, Norwich, UK.

Hull, J., White, A., 1998. Incorporating volatility updating into the historical simulation method for VaR. J. Risk 1, 5–19.

Manganelli, S., 2004. Asset allocation by variance sensitivity analysis. J. Financ. Econom. 2, 370–389.

Patton, A., Sheppard, K., 2009. Evaluating volatility and correlation forecasts. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P., Mikosch, T. (Eds.), Handbook of Financial Time Series. Springer Verlag, Berlin, pp. 801–838.

Pesaran, H., Schleicher, C., Zaffaroni, P., 2009. Model averaging in risk management with an application to futures markets. J. Empir. Finance 16, 280–305.
Pritsker, M., 2006. The hidden dangers of historical simulation. J. Bank. Finance 30, 561–582.
Zaffaroni, P., 2007. Aggregation and memory of models of changing volatility. J. Econom. 136, 237–249.

Empirical Exercises

Open the Chapter8Data.xlsx file from the web site.

1. Construct the 10-day, 1% *VaR* on the last day of the sample using FHS (with 10,000 simulations), RiskMetrics scaling the daily *VaR*s by $\sqrt{10}$ (although it is incorrect), and Monte Carlo simulations of the NGARCH(1,1) model with normally distributed shocks and with parameters as estimated in Chapter 4.
2. Consider counterfactual scenarios where the volatility on the last day of the sample was three times its actual value and also one-half its actual value. Recompute the 10-day *VaR* in exercise 1. What do you see?
3. Repeat exercise 1 computing *ES* rather than *VaR*.
4. Using the DCC model estimated in Chapter 7 try to replicate the correlation forecasts in Figure 8.5, using 10,000 Monte Carlo simulations. Compared with Figure 8.5 do you find evidence of Monte Carlo estimation error when $MC = 10{,}000$?

The answers to these exercises can be found in the Chapter8Results.xlsx file. Which is available in the companion site.

For more information see the companion site at
http://www.elsevierdirect.com/companions/9780123744487

9 Distributions and Copulas for Integrated Risk Management

1 Chapter Overview

In Chapter 7 we considered multivariate risk models that rely on the normal distribution. In Chapter 6 we saw that the univariate normal distribution provides a poor description of asset return distributions—even for well-diversified indexes such as the S&P 500. The normal distribution is convenient but underestimates the probability of large negative returns. The multivariate normal distribution has similar problems. It underestimates the joint probability of simultaneous large negative returns across assets. This in turn means that risk management models built on the multivariate normal distribution are likely to exaggerate the benefits of portfolio diversification. This is clearly not a mistake we want to make as risk managers.

In Chapter 6 we built univariate standardized nonnormal distributions of the shocks

$$z_t \sim D(0, 1)$$

where $z_t = r_t/\sigma_t$ and where $D(*)$ is a standardized univariate distribution.

In this chapter we want to build multivariate distributions for our shocks

$$z_t \sim D(0, \Upsilon_t)$$

where z_t is now a vector of asset specific shocks, $z_{i,t} = r_{i,t}/\sigma_{i,t}$, and where Υ_t is the dynamic correlation matrix. We are assuming that the individual variances have already been modeled using the techniques in Chapters 4 and 5. We are also assuming that the correlation dynamics have been modeled using the DCC model in Chapter 7.

The material in this chapter is relatively complex for two reasons: First, we are departing from the convenient world of normality. Second, we are working with multivariate risk models. The chapter proceeds as follows:

- First, we define and plot threshold correlations, which will be our key graphical tool for detecting multivariate nonnormality.
- Second, we review the multivariate standard normal distribution, and introduce the multivariate standardized symmetric t distribution and the asymmetric extension.

Elements of Financial Risk Management. DOI: 10.1016/B978-0-12-374448-7.00009-9

- Third, we define and develop the copula modeling idea.
- Fourth, we consider risk management and in particular, integrated risk management using the copula model.

2 Threshold Correlations

Just as we used QQ plots to visualize univariate nonnormality in Chapter 6 we need a graphical tool for visualizing nonnormality in the multivariate case. Bivariate threshold correlations are useful in this regard. Consider the daily returns on two assets, for example the S&P 500 and the 10-year bond return introduced in Chapter 7. Threshold correlations are conventional correlations but computed only on a selected subset of the data. Consider a probability p and define the corresponding empirical percentile for asset 1 to be $r_1(p)$ and similarly for asset 2, we have $r_2(p)$. These empirical percentiles, or thresholds, can be viewed as the unconditional *VaR* for each asset. The threshold correlation for probability level p is now defined by

$$\rho(r_{1,t}, r_{2,t}; p) = \begin{cases} Corr\left(r_{1,t}, r_{2,t} | r_{1,t} \leq r_1(p) \text{ and } r_{2,t} \leq r_2(p)\right) & \text{if } p \leq 0.5 \\ Corr\left(r_{1,t}, r_{2,t} | r_{1,t} > r_1(p) \text{ and } r_{2,t} > r_2(p)\right) & \text{if } p > 0.5 \end{cases}$$

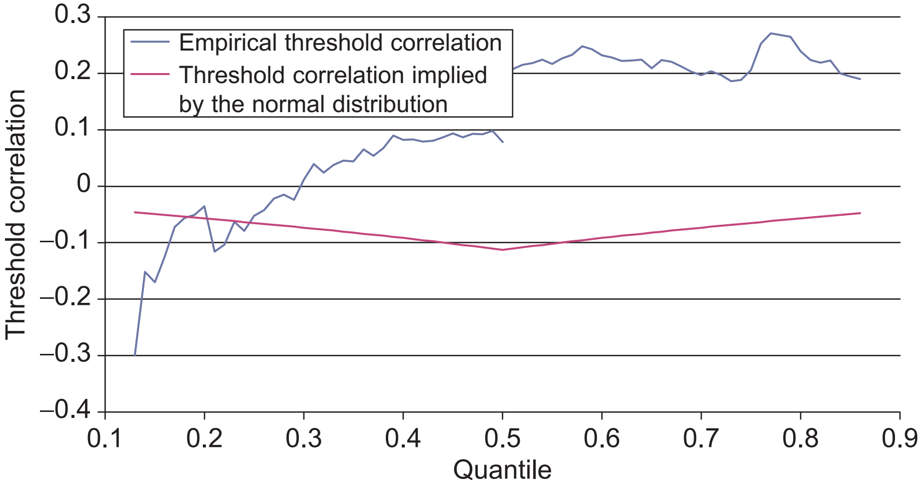
In words, we are computing the correlation between the two assets conditional on both of them being below their pth percentile if $p < 0.5$ and above their pth percentile if $p > 0.5$. In a scatterplot of the two assets we are including only the data in square subsets of the lower-left quadrant when $p < 0.5$ and we are including only the data in square subsets of the upper-right quadrant when $p > 0.5$. If we compute the threshold correlation for a grid of values for p and plot the correlations against p then we get the threshold correlation plot.

The threshold correlations are informative about the dependence across asset returns conditional on both returns being either large and negative or large and positive. They therefore tell us about the tail shape of the bivariate distribution.

The blue line in Figure 9.1 shows the threshold correlation for the S&P 500 return versus the 10-year treasury bond return. When p gets close to 0 or 1 we run out of observations and cannot compute the threshold correlations. We show only correlations where at least 20 observations were available. We use a grid of p values in increments of 0.01. Clearly the most extreme threshold correlations are quite variable and so should perhaps be ignored. Nevertheless, we see an interesting pattern: The threshold correlations get smaller when we observe large negative stock and bond returns simultaneously in the left side of the figure. We also see that large positive stock and bond returns seem to have much higher correlation than the large negative stock and bond returns. This suggests that the bivariate distribution between stock and bond returns is asymmetric.

The red line in Figure 9.1 shows the threshold correlations implied by the bivariate normal distribution when using the average linear correlation coefficient implied by the two return series. Clearly the normal distribution does not match the threshold correlations found in the data.

Figure 9.1 Threshold correlation for S&P 500 versus 10-year treasury bond returns.



Notes: We use daily returns on the S&P 500 index and the 10-year treasury bond index. The blue line shows the threshold correlations from the returns data and the red line shows the threshold correlations implied by the normal distribution with a correlation matching that of the returns data.

Given that we are interested in constructing distributions for the return shocks, rather than the returns themselves we next compute threshold correlations for the shocks as follows:

$$\rho(z_{1,t}, z_{2,t}; p) = \begin{cases} Corr\left(z_{1,t}, z_{2,t} | z_{1,t} \leq z_1(p) \text{ and } z_{2,t} \leq z_2(p)\right) & \text{if } p \leq 0.5 \\ Corr\left(z_{1,t}, z_{2,t} | z_{1,t} > z_1(p) \text{ and } z_{2,t} > z_2(p)\right) & \text{if } p > 0.5 \end{cases}$$

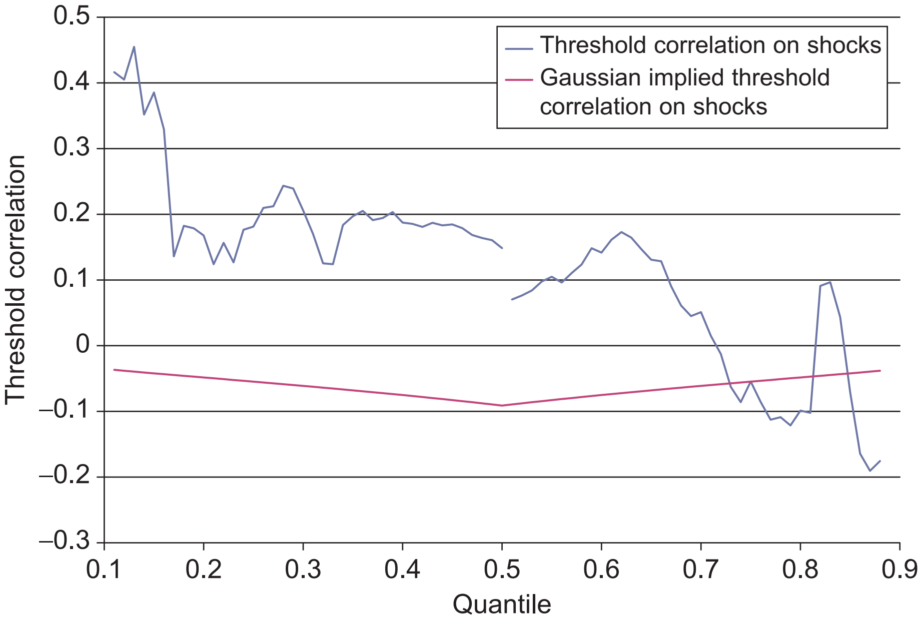
Figure 9.2 shows the threshold correlation plot using the GARCH shocks rather than the returns themselves.

Notice that the patterns are quite different in Figure 9.2 compared with Figure 9.1. Figure 9.2 suggests that the shocks have higher threshold correlations when both shocks are negative than when they are both positive. This indicates that stocks and bonds have important nonlinear left-tail dependencies that risk managers need to model. The threshold correlations implied by the bivariate normal distribution again provide a relatively poor match of the threshold correlations from the empirical shocks.

3 Multivariate Distributions

In this section we consider multivariate distributions that can be combined with GARCH (or RV) and DCC models to provide accurate risk models for large systems of assets. Because we have already modeled the covariance matrix, we need to develop standardized multivariate distributions. We will first review the multivariate standard normal distribution, then we will introduce the multivariate standardized symmetric

Figure 9.2 Threshold correlation for S&P 500 versus 10-year treasury bond GARCH shocks.



Notes: We use daily GARCH shocks on the S&P 500 index and the 10-year treasury bond index. The blue line shows the threshold correlations from the empirical shocks and the red line shows the threshold correlations implied by the normal distribution with a correlation matching that of the empirical shocks.

t distribution, and finally an asymmetric version of the multivariate standardized t distribution.

3.1 The Multivariate Standard Normal Distribution

In Chapter 8 we simulated returns from the normal distribution. In the bivariate case we have the standard normal density with correlation ρ defined by

$$f(z_{1,t}, z_{2,t}; \rho) = \Phi_\rho\left(z_{1,t}, z_{2,t}\right) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{z_{1,t}^2 + z_{2,t}^2 - 2\rho z_{1,t} z_{2,t}}{2(1-\rho^2)}\right)$$

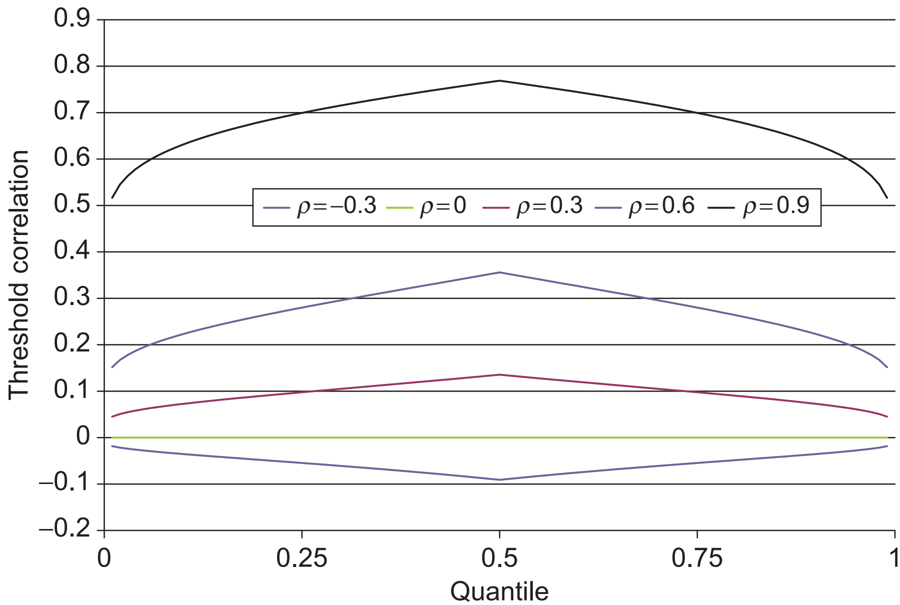
where $1 - \rho^2$ is the determinant of the bivariate correlation matrix

$$|\Upsilon| = \begin{vmatrix} 1 & \rho \\ \rho & 1 \end{vmatrix} = 1 - \rho^2$$

We can of course allow for the correlation ρ to be time varying using the DCC models in Chapter 7.

Figure 9.3 shows the threshold correlation for a bivariate normal distribution for different values of ρ. The figure has been constructed using Monte Carlo random

Figure 9.3 Simulated threshold correlations from bivariate normal distributions with various linear correlations.



Notes: The threshold correlations from the bivariate normal distribution are plotted for various values of the linear correlation parameter.

numbers as in Chapter 8. Notice that regardless of ρ the threshold correlations go to zero as the threshold we consider becomes large (positive or negative). The bivariate normal distribution cannot accurately describe data that has large threshold correlations for extreme values of p.

In the multivariate case with n assets we have the density with correlation matrix Υ

$$f(z_t; \Upsilon) = \Phi_\Upsilon(z_t) = \frac{1}{(2\pi)^{n/2}|\Upsilon|^{1/2}} \exp\left(-\frac{1}{2} z_t' \Upsilon^{-1} z_t\right)$$

which also will have the unfortunate property that each pair of assets in the vector z_t will have threshold correlations that tend to zero for large thresholds. Again we could have a dynamic correlation matrix.

Because of the time-varying variances and correlations we had to use the simulation methods in Chapter 8 to construct multiday *VaR* and *ES*. But we saw in Chapter 7 that the 1-day *VaR* is easily computed via

$$VaR^p_{t+1} = -\sigma_{PF,t+1}\Phi_p^{-1}, \quad \text{where } \sigma_{PF,t+1} = \sqrt{w_t' D_{t+1} \Upsilon_{t+1} D_{t+1} w_t}$$

where we have portfolio weights w_t and the diagonal matrix of standard deviations D_{t+1}.

The 1-day *ES* is also easily computed using

$$ES^p_{t+1} = \sigma_{PF,t+1} \frac{\phi\left(\Phi^{-1}_p\right)}{p}$$

The multivariate normal distribution has the convenient property that a linear combination of multivariate normal variables is also normally distributed. Because a portfolio is nothing more than a linear combination of asset returns, the multivariate normal distribution is very tempting to use. However the fact that it does not adequately capture the (multivariate) risk of returns means that the convenience of the normal distribution comes at a too-high price for risk management purposes. We therefore now consider the multivariate *t* distribution.

3.2 The Multivariate Standardized t Distribution

In Chapter 6 we considered the univariate standardized *t* distribution that had the density

$$f_{\tilde{t}(d)}(z;d) = C(d)\left(1 + z^2/(d-2)\right)^{-(1+d)/2}, \quad \text{for } d > 2$$

where the normalizing constant is

$$C(d) = \frac{\Gamma((d+1)/2)}{\Gamma(d/2)\sqrt{(d-2)\pi}}$$

The bivariate standardized *t* distribution with correlation ρ takes the following form:

$$f_{\tilde{t}(d,\rho)}(z_1, z_2; d, \rho) = C(d,\rho)\left(1 + \frac{z_1^2 + z_2^2 - 2\rho z_1 z_2}{(d-2)\left(1-\rho^2\right)}\right)^{-(d+2)/2}, \quad \text{for } d > 2$$

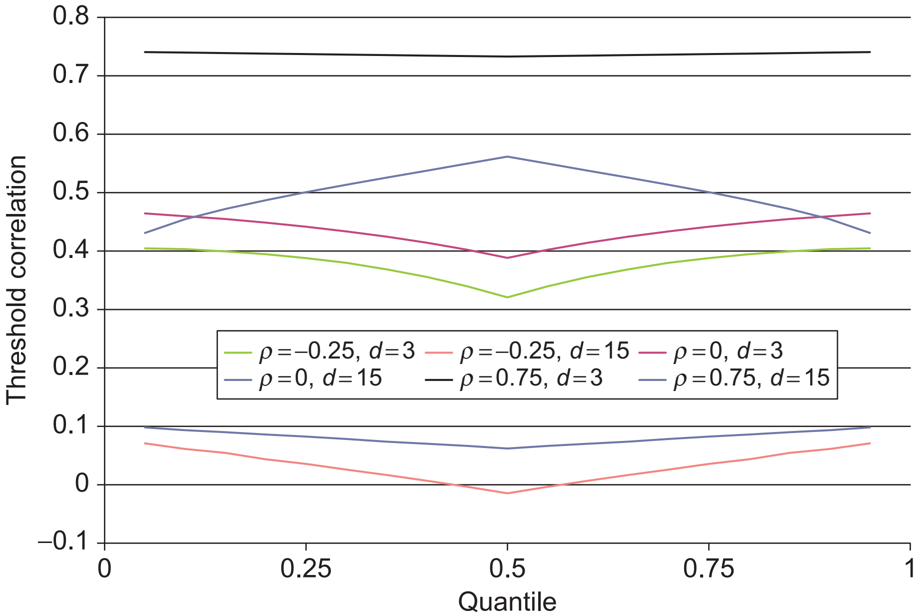
where

$$C(d,\rho) = \frac{\Gamma\left((d+2)/2\right)}{\Gamma(d/2)(d-2)\pi\left(1-\rho^2\right)^{1/2}}$$

Note that d is a scalar here and so the two variables have the same degree of tail fatness.

Figure 9.4 shows simulated threshold correlations of the bivariate standard *t* distribution for different values of d and ρ. Notice that we can generate quite flexible degrees of tail dependence between the two variables when using a multivariate *t* distribution. However, we are constrained in one important sense: Just as the univariate *t* distribution is symmetric so is the multivariate *t* distributions. The threshold correlations are therefore symmetric in the vertical axis.

Figure 9.4 Simulated threshold correlations from the symmetric t distribution with various parameters.



Notes: We simulate a large number of realizations from the bivariate symmetric t distribution. The figure shows the threshold correlations from the simulated data when using various values of the correlation and d parameters.

In the case of n assets we have the multivariate t distribution

$$f_{\tilde{t}(d,\Upsilon)}(z;d,\Upsilon) = C(d,\Upsilon)\left(1+\frac{z'\Upsilon^{-1}z}{(d-2)}\right)^{-(d+n)/2}, \quad \text{for } d>2$$

where

$$C(d,\Upsilon) = \frac{\Gamma((d+n)/2)}{\Gamma(d/2)\,((d-2)\pi)^{n/2}\,|\Upsilon|^{1/2}}$$

Using the density definition we can construct the likelihood function

$$\ln L = \sum_{t=1}^{T} \ln\left(f_{\tilde{t}(d,\Upsilon)}(z_t;d,\Upsilon)\right)$$

which can be maximized to estimate d. The correlation matrix can be preestimated using

$$\Upsilon = \frac{1}{T}\sum_{t=1}^{T} z_t z_t'$$

The correlation matrix Υ can also be made dynamic, which can be estimated in a previous step using the DCC approach in Chapter 7.

Following the logic in Chapter 6, an easier estimate of d can be obtained by computing the kurtosis, ζ_2, of each of the n variables. Recall that the relationship between excess kurtosis and d is

$$\zeta_2 = \frac{6}{d-4}$$

Using all the information in the n variables we can estimate d using

$$d = \frac{6}{\frac{1}{n}\sum_{i=1}^{n}\zeta_{2,i}} + 4$$

where $\zeta_{2,i}$ is the sample excess kurtosis of the ith variable.

A portfolio of multivariate t returns does not itself follow the t distribution unfortunately. We therefore need to rely on Monte Carlo simulation to compute portfolio *VaR* and *ES* even for the 1-day horizon.

The standardized symmetric n dimensional t variable can be simulated as follows:

$$z = \sqrt{\frac{d-2}{d}}\sqrt{W}U$$

where W is a univariate inverse gamma random variable, $W \sim IG\left(\frac{d}{2}, \frac{d}{2}\right)$, and U is a vector of multivariate standard normal variables, $U \sim N(0, \Upsilon)$, and where U and W are independent. This representation can be used to simulate standardized multivariate t variables. First, simulate a scalar random W, then simulate a vector random U (as in Chapter 8), and then construct z as just shown.

The simulated z will have a mean of zero, a standard deviation of one, and a correlation matrix Υ. Once we have simulated MC realizations of the vector z we can use the techniques in Chapter 8 to simulate MC realizations of the vector of asset returns (using GARCH for variances and DCC for correlations), and from this the portfolio *VaR* and *ES* can be computed by simulation as well.

3.3 The Multivariate Asymmetric *t* Distribution

Just as we developed a relatively complex asymmetric univariate t distribution in Chapter 6, we can also develop a relatively complex asymmetric multivariate t distribution.

Let λ be an $n \times 1$ vector of asymmetry parameters. The asymmetric t distribution is then defined by

$$f_{asy\tilde{t}}(z; d, \lambda, \Upsilon) = \frac{C_{asy}\left(d, \dot{\Upsilon}\right) K_{\frac{d+n}{2}}\left(\sqrt{\left(d + (z-\dot{\mu})'\,\dot{\Upsilon}^{-1}(z-\dot{\mu})\right)\lambda'\dot{\Upsilon}^{-1}\lambda}\right)\left(1 + \frac{1}{d}(z-\dot{\mu})'\,\dot{\Upsilon}^{-1}(z-\dot{\mu})\right)^{-(d+n)/2}}{\exp\left(-(z-\dot{\mu})'\,\dot{\Upsilon}^{-1}\lambda\right)\left(\sqrt{\left(d + (z-\dot{\mu})'\,\dot{\Upsilon}^{-1}(z-\dot{\mu})\right)\lambda'\dot{\Upsilon}^{-1}\lambda}\right)^{-(d+n)/2}}$$

where

$$\dot{\mu} = -\frac{d}{d-2}\lambda,$$
$$\dot{\Upsilon} = \frac{d-2}{d}\left(\Upsilon - \frac{2d^2}{(d-2)^2(d-4)}\lambda\lambda'\right), \text{ and}$$
$$C_{asy}\left(d, \dot{\Upsilon}\right) = \frac{2^{(1-(d+n)/2)}}{\Gamma(d/2)\,(d\pi)^{n/2}\,|\dot{\Upsilon}|^{\frac{1}{2}}}$$

and where $K_{\frac{d+n}{2}}(x)$ is the so-called modified Bessel function of the third kind, which can be evaluated in Excel using the formula besselk(x, $(d+n)/2$).

Note that the vector $\dot{\mu}$ and matrix $\dot{\Upsilon}$ are constructed so that the vector of random shocks z will have a mean of zero, a standard deviation of one, and the correlation matrix Υ. Note also that if $\lambda = 0$ then $\dot{\mu} = 0$ and $\dot{\Upsilon} = \frac{d-2}{d}\Upsilon$.

Although it is not obvious from this definition of $f_{asy\tilde{t}}(z; d, \lambda)$, we can show that the asymmetric t distribution will converge to the symmetric t distribution as the asymmetry parameter vector λ goes to a vector of zeros.

Figure 9.5 shows simulated threshold correlations of the bivariate asymmetric t distribution when setting $\lambda = 0.2$ for both assets, $d = 10$, and when considering different values of ρ. Look closely at Figure 9.5. Note that the asymmetric t distribution is able to capture asymmetries in the threshold correlations and gaps in the threshold correlation around the median (the 0.5 quantile on the horizontal axis), which we saw in the stock and bond thresholds in Figures 9.1 and 9.2.

From the density $f_{asy\tilde{t}}(z; d, \lambda, \Upsilon)$ we can construct the likelihood function

$$\ln L = \sum_{t=1}^{T} \ln\left(f_{asy\tilde{t}}(z_t; d, \lambda, \Upsilon)\right)$$

which can be maximized to estimate the scalar d and and vector λ. As before, the correlation matrix can be preestimated using

$$\Upsilon = \frac{1}{T}\sum_{t=1}^{T} z_t z_t'$$

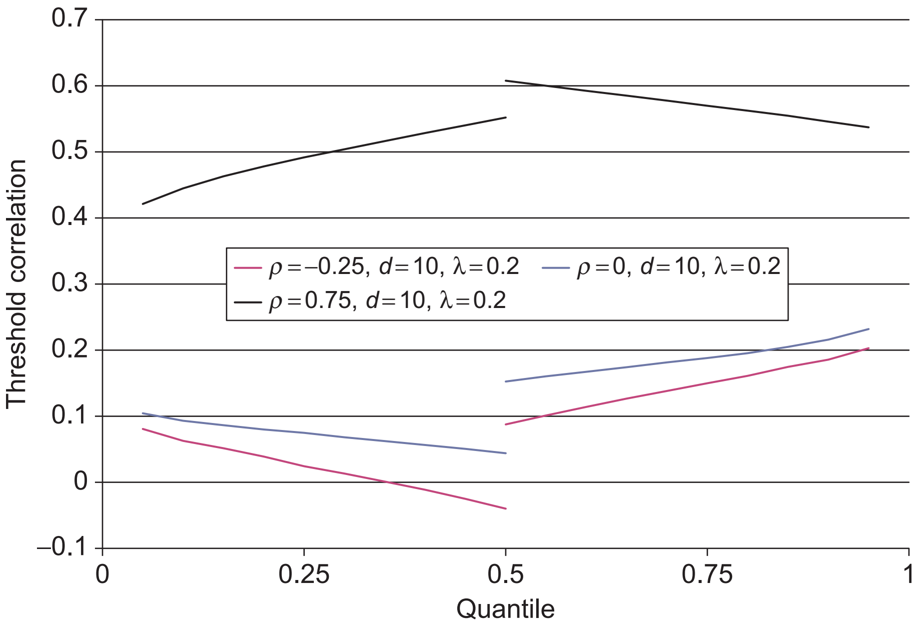
The correlation matrix Υ can also be made dynamic, Υ_t, which can be estimated in a previous step using the DCC approach in Chapter 7 as mentioned earlier.

Simulated values of the (nonstandardized) asymmetric t distribution can be constructed from inverse gamma and normal variables. We now have

$$z = \dot{\mu} + \sqrt{W}U + \lambda W$$

where W is again an inverse gamma variable $W \sim IG\left(\frac{d}{2}, \frac{d}{2}\right)$, U is a vector of normal variables, $U \sim N\left(0, \dot{\Upsilon}\right)$, and U and W are independent. Note that the asymmetric t distribution generalizes the symmetric t distribution by adding a term related to the same inverse gamma random variable W, which is now scaled by the asymmetry vector λ.

Figure 9.5 Simulated threshold correlations from the asymmetric t distribution with various linear correlations.



Notes: We simulate a large number of realizations from the bivariate asymmetric t distribution. The figure shows the threshold correlations from the simulated data when using various correlation values.

The simulated z vector will have the following mean:

$$E[z] = \dot{\mu} + \frac{d}{d-2}\lambda = 0$$

where we have used the definition of $\dot{\mu}$ from before. The variance-covariance matrix of the simulated shocks will be

$$Cov(z) = \frac{d}{d-2}\dot{\Upsilon} + \frac{2d^2}{(d-2)^2(d-4)}\lambda\lambda' = \Upsilon$$

where we have used the definition of $\dot{\Upsilon}$ from before.

The asymmetric t distribution allows for much more flexibility than the symmetric t distribution because of the vector of asymmetry parameters, λ. However in large dimensions (i.e., for a large number of assets, n) estimating the n different λs may be difficult.

Note that the scalar d and the vector λ have to describe the n univariate distributions as well as the joint density of the n assets. We may be able to generate even more flexibility by modeling the univariate distributions separately using for example the

asymmetric t distribution in Chapter 6. In this case each asset i would have its own $d_{1,i}$ and its own $d_{2,i}$ (using Chapter 6 notation) capturing univariate skewness and kurtosis. But we then need a method for linking the n distributions together. Fortunately, this is exactly what copula models do.

4 The Copula Modeling Approach

The multivariate normal distribution underestimates the threshold correlations typically found in daily returns. The multivariate t distribution allows for larger threshold correlations but the condition that the d parameter is the same across all assets is restrictive. The asymmetric t distribution is more flexible but it requires estimating many parameters simultaneously.

Ideally we would like to have a modeling approach where the univariate models from Chapters 4 through 6 can be combined to form a proper multivariate distribution. Fortunately, the so-called copula functions have been developed in statistics to provide us exactly with the tool we need.

Consider n assets with potentially different univariate (also known as marginal) distributions, $f_i(z_i)$ and cumulative density functions (CDFs) $u_i = F_i(z_i)$ for $i = 1, 2, \ldots, n$. Note that u_i is simply the probability of observing a value below z_i for asset i. Our goal is to link the marginal distributions across the assets to generate a valid multivariate density.

4.1 Sklar's Theorem

Sklar's theorem provides us with the theoretical foundation we need. It states that for a very general class of multivariate cumulative density functions, defined as $F(z_1, \ldots, z_n)$, with marginal CDFs $F_1(z_1), \ldots, F_n(z_n)$, there exists a unique copula function, $G(\bullet)$ linking the marginals to form the joint distribution

$$\begin{aligned} F(z_1, \ldots, z_n) &= G(F_1(z_1), \ldots, F_n(z_n)) \\ &= G(u_1, \ldots, u_n) \end{aligned}$$

The $G(u_1, \ldots, u_n)$ function is sometimes known as the copula CDF.

Sklar's theorem then implies that the multivariate probability density function (PDF) is

$$\begin{aligned} f(z_1, \ldots, z_n) &= \frac{\partial^n G(F_1(z_1), \ldots, F_n(z_n))}{\partial z_1 \cdots \partial z_n} \\ &= \frac{\partial^n G(u_1, \ldots, u_n)}{\partial u_1 \cdots \partial u_n} \times \prod_{i=1}^{n} f_i(z_i) \\ &= g(u_1, \ldots, u_n) \times \prod_{i=1}^{n} f_i(z_i) \end{aligned}$$

where the copula PDF is defined in the last equation as

$$g(u_1, \ldots, u_n) \equiv \frac{\partial^n G(u_1, \ldots, u_n)}{\partial u_1 \cdots \partial u_n}$$

Consider now the logarithm of the PDF

$$\ln f(z_1, \ldots, z_n) = \ln g(u_1, \ldots, u_n) + \sum_{i=1}^{n} \ln f_i(z_i)$$

This decomposition shows that we can build the large and complex multivariate density in a number of much easier steps: First, we build and estimate n potentially different marginal distribution models $f_1(z_1), \ldots, f_n(z_n)$ using the methods in Chapters 4 through 6. Second, we decide on the copula PDF $g(u_1, \ldots, u_n)$ and estimate it using the probability outputs u_i from the marginals as the data.

Notice how Sklar's theorem offers a very powerful framework for risk model builders. Notice also the analogy with GARCH and DCC model building: The DCC correlation model allows us to use different GARCH models for each asset. Similarly copula models allow us to use a different univariate density model for each asset.

The log likelihood function corresponding to the entire copula distribution model is constructed by summing the log PDF over the T observations in our sample

$$\ln L = \sum_{t=1}^{T} \ln g(u_{1,t}, \ldots, u_{n,t}) + \sum_{t=1}^{T} \sum_{i=1}^{n} \ln f_i(z_{i,t})$$

But if we have estimated the n marginal distributions in a first step then the copula likelihood function is simply

$$\ln L_g = \sum_{t=1}^{T} \ln g(u_{1,t}, \ldots, u_{n,t})$$

The upshot of this is that we only have to estimate the parameters in the copula PDF function $g(u_{1,t}, \ldots, u_{n,t})$ in a single step. We can estimate all the parameters in the marginal PDFs beforehand. This makes high-dimensional modeling possible. We can for example allow for each asset to follow different univariate asymmetric t distributions (from Chapter 6) each estimated one at a time. Taking these asset-specific distributions as given we can then link them together by estimating the parameters in $g(u_{1,t}, \ldots, u_{n,t})$ in the second step.

Sklar's theorem is very general: It holds for a large class of multivariate distributions. However it is not very specific: It does not say anything about the functional form of $G(u_1, \ldots, u_n)$ and thus $g(u_{1,t}, \ldots, u_{n,t})$. In order to implement the copula modelling approach we need to make specific modeling choices for the copula CDF.

4.2 The Normal Copula

After Sklar's theorem was published in 1959 researchers began to search for potential specific forms for the copula function. Given that the copula CDF must take as inputs marginal CDFs and deliver as output a multivariate CDF one line of research simply took known multivariate distributions and reverse engineered them to take as input probabilities, u, instead of shocks, z.

The most convenient multivariate distribution is the standard normal, and from this we can build the normal copula function. In the bivariate case we have

$$\begin{aligned} G(u_1, u_2; \rho^*) &= \Phi_{\rho^*}(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \\ &= \Phi_{\rho^*}(\Phi^{-1}(F_1(z_1)), \Phi^{-1}(F_2(z_2))) \end{aligned}$$

where ρ^* is the correlation between $\Phi^{-1}(u_1)$ and $\Phi^{-1}(u_2)$ and we will refer to it as the copula correlation. As in previous chapters, $\Phi^{-1}(\bullet)$ denotes the univariate standard normal inverse CDF.

Note that if the two marginal densities, F_1 and F_2, are standard normal then we get

$$\begin{aligned} G(u_1, u_2; \rho^*) &= \Phi_{\rho^*}(\Phi^{-1}(\Phi(z_1)), \Phi^{-1}(\Phi(z_2))) \\ &= \Phi_{\rho^*}(z_1, z_2) \end{aligned}$$

which is simply the bivariate normal distribution. But note also that if the marginal distributions are NOT the normal then the normal copula does NOT imply the normal distribution. The normal copula is much more flexible than the normal distribution because the normal copula allows for the marginals to be nonnormal, which in turn can generate a multitude of nonnormal multivariate distributions.

In order to estimate the normal copula we need the normal copula PDF. It can be derived as

$$\begin{aligned} g(u_1, u_2; \rho^*) &= \frac{\phi_{\rho^*}(\Phi^{-1}(u_1), \Phi^{-1}(u_2))}{\phi\left(\Phi^{-1}(u_1)\right)\phi\left(\Phi^{-1}(u_2)\right)} \\ &= \frac{1}{\sqrt{1-\rho^{*2}}} \exp\left\{ -\frac{\Phi^{-1}(u_1)^2 + \Phi^{-1}(u_2)^2 - 2\rho^*\Phi^{-1}(u_1)\Phi^{-1}(u_2)}{2(1-\rho^{*2})} \right. \\ &\qquad \left. + \frac{\Phi^{-1}(u_1)^2 + \Phi^{-1}(u_2)^2}{2} \right\} \end{aligned}$$

where $\phi_{\rho^*}(*)$ denotes the bivariate standard normal PDF and $\phi(*)$ denotes the univariate standard normal PDF. The copula correlation, ρ^*, can now be estimated by

maximizing the likelihood

$$\ln L_g = \sum_{t=1}^{T} \ln g(u_{1,t}, u_{2,t}) = -\frac{T}{2}\ln\left(1-\rho^{*2}\right)$$
$$-\sum_{t=1}^{T} \frac{\Phi^{-1}(u_{1,t})^2 + \Phi^{-1}(u_{2,t})^2 - 2\rho^*\Phi^{-1}(u_{1,t})\Phi^{-1}(u_{2,t})}{2(1-\rho^{*2})}$$
$$+\frac{\Phi^{-1}(u_{1,t})^2 + \Phi^{-1}(u_{2,t})^2}{2}$$

where we have $u_{1,t} = F_1\left(z_{1,t}\right)$ and $u_{2,t} = F_2\left(z_{2,t}\right)$.

In the general case with n assets we have the multivariate normal copula CDF and copula PDF

$$G(u_1,\ldots,u_n;\Upsilon^*) = \Phi_{\Upsilon^*}(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_n))$$
$$g(u_1,\ldots,u_2;\Upsilon^*) = \frac{\phi_{\Upsilon^*}(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_n))}{\prod_{i=1}^{n}\phi\left(\Phi^{-1}(u_i)\right)}$$
$$= \left|\Upsilon^*\right|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\Phi^{-1}(u)'(\Upsilon^{*-1}-I_n)\Phi^{-1}(u)\right\}$$

where u is the vector with elements $(u_1,\ldots,u_n)$, and where I_n is an n-dimensional identity matrix that has ones on the diagonal and zeros elsewhere. The correlation matrix, Υ^*, in the normal copula can be estimated by maximizing the likelihood

$$\ln L_g = \sum_{t=1}^{T}\ln g(u_{1,t},\ldots,u_{n,t})$$
$$= -\frac{1}{2}\sum_{t=1}^{T}\ln\left|\Upsilon^*\right| - \frac{1}{2}\sum_{t=1}^{T}\Phi^{-1}(u_t)'(\Upsilon^{*-1}-I_n)\Phi^{-1}(u_t)$$

If the number of assets is large then Υ^* contains many elements to be estimated and numerical optimization will be difficult.

Let us define the copula shocks for asset i on day t as follows:

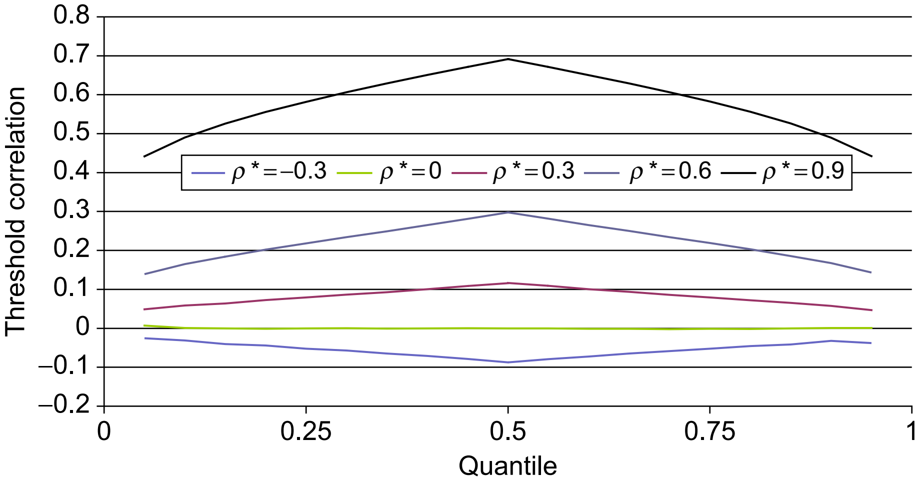
$$z_{i,t}^* = \Phi^{-1}(u_{i,t}) = \Phi^{-1}\left(F_i(z_{i,t})\right)$$

An estimate of the copula correlation matrix can be obtained via correlation targeting

$$\Upsilon^* = \frac{1}{T}\sum_{t=1}^{T} z_t^* z_t^{*\prime}$$

In small dimensions this can be used as starting values of the MLE optimization. In large dimensions it provides a feasible estimate where the MLE is infeasible.

Figure 9.6 Simulated threshold correlations from the bivariate normal copula with various copula correlations.



Notes: We simulate a large number of realizations from the bivariate normal copula. The figure shows the threshold correlations from the simulated data when using various values of the copula correlation parameter.

Consider again the previous bivariate normal copula. We have the bivariate distribution

$$
\begin{aligned}
F(z_1, z_2) &= G(u_1, u_2) \\
&= \Phi_{\rho^*}(\Phi^{-1}(u_1), \Phi^{-1}(u_2))
\end{aligned}
$$

Figure 9.6 shows the threshold correlation between u_1 and u_2 for different values of the copula correlation ρ^*. Naturally, the normal copula threshold correlations look similar to the normal distribution threshold correlations in Figure 9.3.

Note that the threshold correlations are computed from the u_1 and u_2 probabilities and not from the z_1 and z_2 shocks, which was the case in Figures 9.1 through 9.5. The normal copula gives us flexibility by allowing the marginal distributions F_1 and F_2 to be flexible but the multivariate aspects of the normal distribution remains: The threshold correlations go to zero for extreme u_1 and u_2 observations, which is likely not desirable in a risk management model where extreme moves are often highly correlated across assets.

4.3 The t Copula

The normal copula is relatively convenient and much more flexible than the normal distribution but for many financial risk applications it does not allow for enough dependence between the tails of the distributions of the different assets. This was

illustrated by the normal copula threshold correlations in Figure 9.6, which decay to zero for extreme tails.

Fortunately a copula model can be built from the t distribution as well. Consider first the bivariate case. The bivariate t copula CDF is defined by

$$G(u_1, u_2; \rho^*, d) = t_{(d,\rho^*)}\left(t^{-1}(u_1; d), t^{-1}(u_2; d)\right)$$

where $t_{(d,\rho^*)}(*)$ denotes the (not standardized) symmetric multivariate t distribution, and $t^{-1}(u; d)$ denotes the inverse CDF of the symmetric (not standardized) univariate t distribution, which we denoted $t_{u_1}^{-1}(d)$ in Chapter 6.

The corresponding bivariate t copula PDF is

$$\begin{aligned} g(u_1, u_2; \rho^*, d) &= \frac{t_{(d,\rho^*)}(t^{-1}(u_1; d), t^{-1}(u_2; d))}{f_{t(d)}\left(t^{-1}(u_1; d); d\right) f_{t(d)}\left(t^{-1}(u_2; d); d\right)} \\ &= \frac{\Gamma\left(\frac{d+2}{2}\right)}{\sqrt{1-\rho^2}\,\Gamma\left(\frac{d}{2}\right)} \left(\frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d+1}{2}\right)}\right)^2 \\ &\times \frac{\left(1 + \frac{(t^{-1}(u_1; d))^2 + (t^{-1}(u_2; d))^2 - 2\rho t^{-1}(u_1; d) t^{-1}(u_1; d)}{d(1-\rho^2)}\right)^{-\frac{d+2}{2}}}{\left(1 + \frac{(t^{-1}(u_1; d))^2}{d}\right)^{-\frac{d+1}{2}} \left(1 + \frac{(t^{-1}(u_2; d))^2}{d}\right)^{-\frac{d+1}{2}}} \end{aligned}$$

In Figure 9.7 we plot the threshold correlation between u_1 and u_2 for different values of the copula correlation ρ^* and the tail fatness parameter d. Naturally, the t copula threshold correlations look similar to the t distribution threshold correlations in Figure 9.4 but different from the normal threshold correlations in Figure 9.6.

The t copula can generate large threshold correlations for extreme moves in the assets. Furthermore it allows for individual modeling of the marginal distributions, which allows for much flexibility in the resulting multivariate distribution.

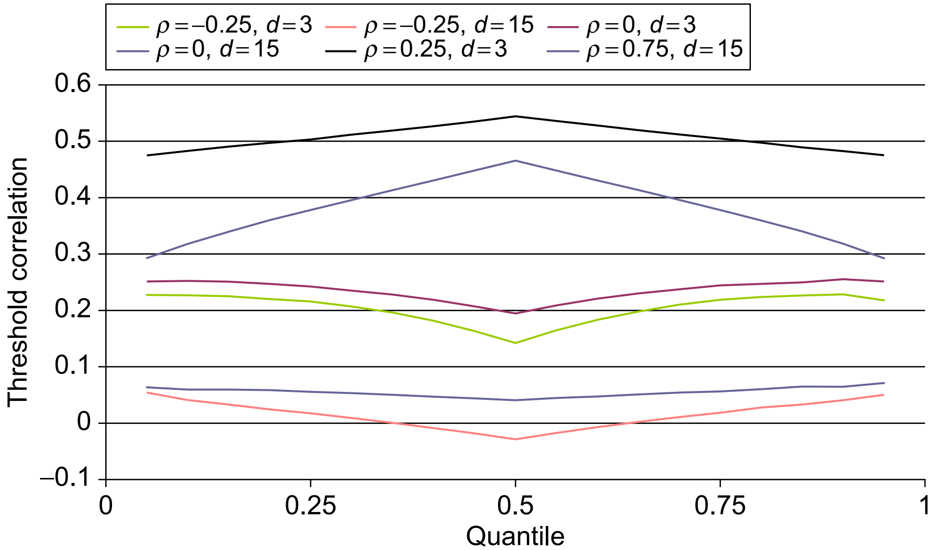
In the general case of n assets we have the t copula CDF

$$G(u_1, \ldots, u_n; \Upsilon^*, d) = t_{(d,\Upsilon^*)}\left(t^{-1}(u_1; d), \ldots, t^{-1}(u_n; d)\right)$$

and the t copula PDF

$$\begin{aligned} g(u_1, \ldots, u_n; \Upsilon^*, d) &= \frac{t_{(d,\Upsilon^*)}(t^{-1}(u_1; d), \ldots, t^{-1}(u_n; d))}{\prod_{i=1}^n t\left(t^{-1}(u_i; d); d\right)} \\ &= \frac{\Gamma\left(\frac{d+n}{2}\right)}{|\Upsilon^*|^{\frac{1}{2}}\Gamma\left(\frac{d}{2}\right)} \left(\frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d+1}{2}\right)}\right)^n \frac{\left(1 + \frac{1}{d} t^{-1}(u; d)' \Upsilon^{*-1} t^{-1}(u; d)\right)^{-\frac{d+n}{2}}}{\prod_{i=1}^n \left(1 + \frac{(t^{-1}(u_i; d))^2}{d}\right)^{-\frac{d+1}{2}}} \end{aligned}$$

Figure 9.7 Simulated threshold correlations from the symmetric t copula with various parameters.



Notes: We simulate a large number of realizations from the bivariate symmetric t copula. The figure shows the threshold correlations from the simulated data when using various values of the copula correlation and d parameter.

Notice that d is a scalar, which makes the t copula somewhat restrictive but also makes it implementable for many assets.

Maximum likelihood estimation can again be used to estimate the parameters d and Υ^* in the t copula. We need to maximize

$$\ln L_g = \sum_{t=1}^{T} \ln g(u_{1,t}, \ldots, u_{n,t})$$

defining again the copula shocks for asset i on day t as follows:

$$z_{i,t}^* = t^{-1}(u_{i,t}; d) = t^{-1}\left(F_i\left(z_{i,t}\right); d\right)$$

In large dimensions we need to target the copula correlation matrix, which can be done as before using

$$\Upsilon^* = \frac{1}{T}\sum_{t=1}^{T} z_t^* z_t^{*\prime}$$

With this matrix preestimated we will only be searching for the parameter d in the maximization of $\ln L_g$ earlier.

4.4 Other Copula Models

An asymmetric t copula can be developed from the asymmetric multivariate t distribution in the same way that we developed the symmetric t copula from the multivariate t distribution earlier.

Figure 9.8 shows the iso-probability or probability contour plots of the bivariate normal copula, the symmetric t copula, and the asymmetric (or skewed) t copula with positive or negative λ. Each line in the contour plot represents the combinations of z_1 and z_2 that correspond to an equal level of probability. The more extreme values of z_1 and z_2 in the outer contours therefore correspond to lower levels of probability. We have essentially taken the bivariate distribution, which is a 3D graph, and sliced it at different levels of probability. The probability levels for each ring are the same across the four panels in Figure 9.8.

Consider the bottom-left corner of each panel in Figure 9.8. This corresponds to extreme outcomes where both assets have a large negative shock. Notice that the symmetric t copula and particularly the asymmetric t copula with negative λ can accommodate the largest (negative) shocks on the outer contours. The two univariate distributions are assumed to be standard normal in Figure 9.8.

In large dimensions it may be necessary to restrict the asymmetry parameter λ to be the same across all or across subsets of the assets. But note that the asymmetric t copula still offers flexibility because we can use the univariate asymmetric t distribution in Chapter 6 to model the marginal distributions so that the λ in the asymmetric t copula only has to capture multivariate aspects of asymmetry. In the multivariate asymmetric t distribution the vector of λ parameters needs to capture asset-specific as well as multivariate asymmetries.

We have only considered normal and t copulas here. Other classes of copula functions exist as well. However, only a few copula functions are applicable in high dimensions; that is, when the number of assets, n, is large.

So far we have assumed that the copula correlation matrix, Υ^*, is constant across time. However, we can let the copula correlations be dynamic using the DCC approach in Chapter 7. We would now use the copula shocks $z_{i,t}^*$ as data input into the estimation of the dynamic copula correlations instead of the $z_{i,t}$ that were used in Chapter 7.

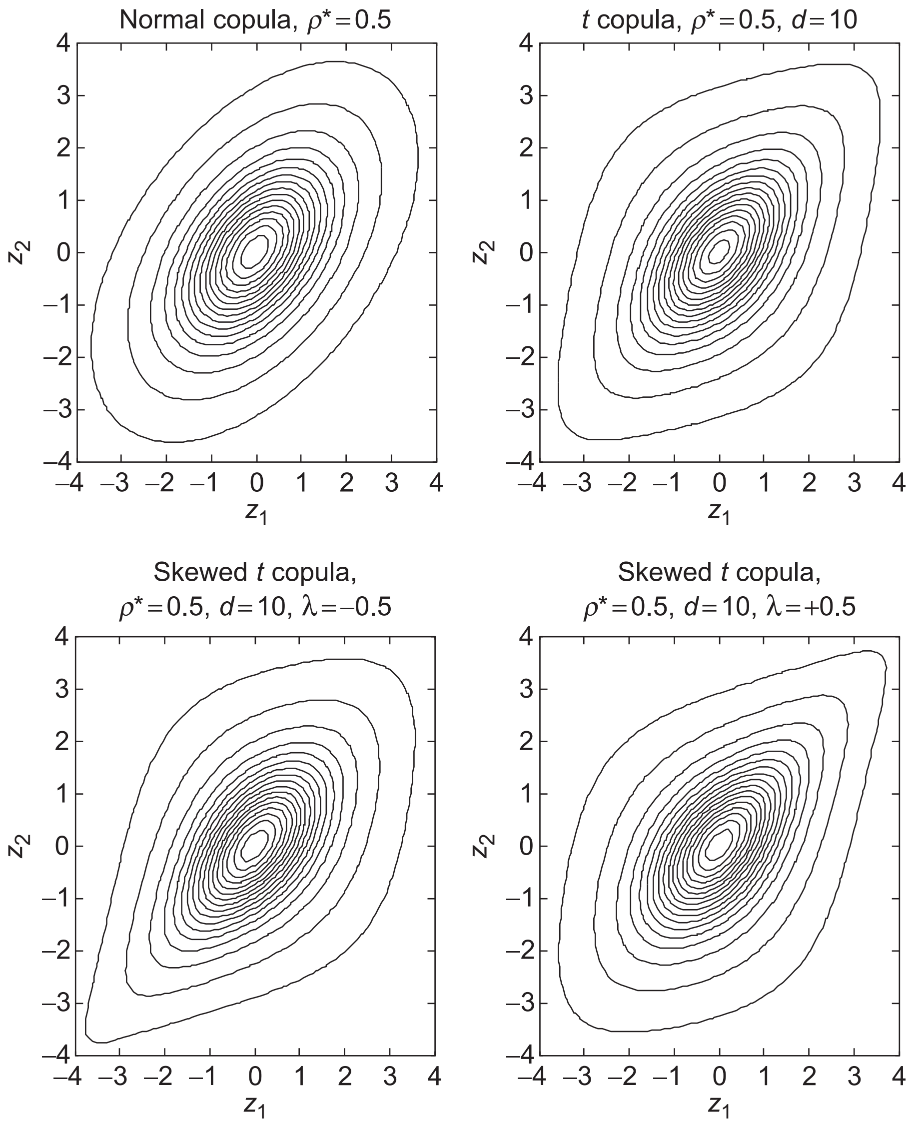
5 Risk Management Using Copula Models

5.1 Copula VaR and ES by Simulation

When we want to compute portfolio *VaR* and *ES* from copula models we need to rely on Monte Carlo simulation. Monte Carlo simulation essentially reverses the steps taken in model building. Recall that we have built the copula model from returns as follows:

- First, estimate a dynamic volatility model, $\sigma_{i,t}$ (Chapters 4 and 5), on each asset to get from observed return $R_{i,t}$ to shock $z_{i,t} = r_{i,t}/\sigma_{i,t}$.

Figure 9.8 Contour probability plots for the normal, symmetric *t*, and asymmetric skewed *t* copula.



Notes: We plot the contour probabilities for the normal, symmetric *t*, and asymmetric skewed *t* copulas. The marginal distributions are assumed to be standard normal. Each line on the figure corresponds to a particular probability level. The probability levels are held fixed across the four panels.

- Second, estimate a density model for each asset (Chapter 6) to get the probabilities $u_{i,t} = F_i(z_{i,t})$ for each asset.
- Third, estimate the parameters in the copula model using $\ln L_g = \sum_{t=1}^{T} \ln g(u_{1,t}, \ldots, u_{n,t})$.

When we simulate data from the copula model we need to reverse the steps taken in the estimation of the model. We get the algorithm:

- First, simulate the probabilities $(u_{1,t}, \ldots, u_{n,t})$ from the copula model.
- Second, create shocks from the copula probabilities using the marginal inverse CDFs $z_{i,t} = F_i^{-1}(u_{i,t})$ on each asset.
- Third, create returns from shocks using the dynamic volatility models, $r_{i,t} = \sigma_{i,t} z_{i,t}$ on each asset.

Once we have simulated MC vectors of returns from the model we can easily compute the simulated portfolio returns using a given portfolio allocation. The portfolio *VaR*, *ES*, and other measures can then be computed on the simulated portfolio returns in Chapter 8. For example, the 1% *VaR* will be the first percentile of all the simulated portfolio return paths.

5.2 Integrated Risk Management

Integrated risk management is concerned with the aggregation of risks across different business units within an organization. Each business unit may have its own risk model but the senior management needs to know the overall risk to the organization arising in the aggregate from the different units. In short, senior management needs a method for combining the marginal distributions of returns in each business unit.

In the simplest (but highly unrealistic) case, we can assume that the multivariate normal model gives a good description of the overall risk of the firm. If the correlations between all the units are one (also not realistic) then we get a very simple result. Consider first the bivariate case

$$\begin{aligned} VaR_{t+1}^{p} &= -\sqrt{w_{1,t}^2\sigma_{1,t}^2 + w_{2,t}^2\sigma_{2,t}^2 + 2w_{1,t}w_{2,t}\rho_{12,t}\sigma_{1,t}\sigma_{2,t}}\,\Phi_p^{-1} \\ &= -\sqrt{\left(w_{1,t}\sigma_{1,t} + w_{2,t}\sigma_{2,t}\right)^2}\,\Phi_p^{-1} \\ &= \left(w_{1,t}VaR_{1,t+1}^{p} + w_{2,t}VaR_{2,t+1}^{p}\right) \end{aligned}$$

where we have assumed the weights are positive. The total *VaR* is simply the (weighted) sum of the two individual business unit *VaR*s under these specific assumptions.

In the general case of n business units we similarly have

$$VaR_{t+1}^{p} = \sum_{i=1}^{n} w_{i,t} VaR_{i,t+1}^{p}$$

but again only when the returns are multivariate normal with correlation equal to one between all pairs of units.

In the more general case where the returns are not normally distributed with all correlations equal to one, we need to specify the multivariate distribution from the individual risk models. Copulas do exactly that and they are therefore very well suited for integrated risk management. But we do need to estimate the copula parameters and also need to rely on Monte Carlo simulation to compute organization wide *VaR*s and other risk measures. The methods in this and the previous chapter can be used for this purpose.

6 Summary

Multivariate risk models require assumptions about the multivariate distribution of return shocks. The multivariate normal distribution is by far the most convenient model but it does not allow for enough extreme dependence in most risk management applications. We can use the threshold correlation to measure extreme dependence in observed asset returns and in the available multivariate distributions. The multivariate symmetric t and in particular the asymmetric t distribution provides the larger threshold correlations that we need, but in high dimension the asymmetric t may be cumbersome to estimate. Copula models allow us to link together a wide range of marginal distributions. The normal and t copulas we have studied are fairly flexible and are applicable in high dimensions. Copulas are also well suited for integrated risk management where the risk models from individual business units must be linked together to provide a sensible aggregate measure of risk for the organization as a whole.

Further Resources

For powerful applications of threshold correlations in equity markets, see Longin and Solnik (2001), Ang and Chen (2002), and Okimoto (2008).

Sklar's theorem is proved in Sklar (1959). The multivariate symmetric and asymmetric t distributions are analyzed in Demarta and McNeil (2005), who also develop the t copula model. Jondeau and Rockinger (2006) develop the copula-GARCH approach advocated here.

Thorough treatments of copula models are provided in the books by Cherubini et al. (2004) and McNeil et al. (2005). Surveys focusing on risk management applications of copulas can be found in Embrechts et al. (2003, 2002), Fischer et al. (2009), and Patton (2009).

Model selection in the context of copulas is studied in Chen and Fan (2006) and Kole et al. (2007). Default correlation modeling using copulas is done in Li (2000).

Dynamic copula models have been developed in Patton (2004, 2006), Patton and Oh (2011), Chollete et al. (2009), Christoffersen et al. (2011), Christoffersen and Langlois (2011), and Creal et al. (2011). Hafner and Manner (2010) suggest a stochastic copula approach that requires simulation in estimation.

A framework for integrated risk management using copulas is developed in Rosenberg and Schuermann (2006). Copula models are also well suited for studying financial contagion as done in Rodriguez (2007).

References

Ang, A., Chen, J., 2002. Asymmetric correlations of equity portfolios. J. Financ. Econ. 63, 443–494.

Chen, X., Fan, Y., 2006. Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. J. Econom. 135, 125–154.

Cherubini, U., Luciano, E., Vecchiato, E., 2004. Copula Methods in Finance. Wiley, New York.

Chollete, L., Heinen, A., Valdesogo, A., 2009. Modeling international financial returns with a multivariate regime-switching copula. J. Financ. Econom. 7, 437–480.

Christoffersen, P., Errunza, V., Jacobs, K., Langlois, H., 2011. Is the potential for international diversification disappearing? Available from: SSRN, http://ssrn.com/abstract=1573345.

Christoffersen, P., Langlois, H., 2011. The joint dynamics of equity market factors. Working Paper, University of Toronto.

Creal, D., Koopman, S., Lucas, A., 2011. A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. J. Bus. Econ. Stat. forthcoming.

Demarta, S., McNeil, A., 2005. The *t* copula and related copulas. Int. Stat. Rev. 73, 111–129.

Embrechts, P., Lindskog, F., McNeil, A., 2003. Modelling dependence with copulas and applications to risk management. In: Rachev, S. (Ed.), Handbook of Heavy Tailed Distributions in Finance. Elsevier, Amsterdam, The Netherlands, Chapter 8, pp. 329–384.

Embrechts, P., McNeil, A., Straumann, D., 2002. Correlation and dependence in risk management: Properties and pitfalls. In: Dempster, M.A.H. (Ed.), Risk Management: Value at Risk and Beyond. Cambridge University Press, pp. 176–223.

Fischer, M., Kock, C., Schluter, S., Weigert, F., 2009. An empirical analysis of multivariate copula models. Quant. Finance 9, 839–854.

Hafner, C.M., Manner, H., 2010. Dynamic stochastic copula models: Estimation, inference and application. J. Appl. Econom. forthcoming.

Jondeau, R. Rockinger, M., 2006. The copula-GARCH model of conditional dependencies: An international stock market application. J. Int. Money Finance 25, 827–853.

Kole, E., Koedijk, K., Verbeek, M., 2007. Selecting copulas for risk management. J. Bank. Finance 31, 2405–2423.

Li, D.X., 2000. On default correlation: A copula function approach. J. Fixed Income 9, 43–54.

Longin, F., Solnik, B., 2001. Extreme correlation of international equity markets. J Finance. 56, 651–678.

McNeil, A., Frey, R., Embrechts, P., 2005. Quantitative Risk Management: Concepts, Techniques, and Tools. Princeton University Press, Princeton, NJ.

Okimoto, T. 2008. New evidence of asymmetric dependence structures in international equity markets. J. Financ. Quant. Anal. 43, 787–815.

Patton, A., 2004. On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. J. Financ. Econom. 2, 130–168.

Patton, A., 2006. Modeling asymmetric exchange rate dependence. Int. Econ. Rev. 47, 527–556.

Patton, A., 2009. Copula-based models for financial time series. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P., Mikosch, T. (Eds.), Handbook of Financial Time Series. Springer Verlag, Berlin.

Patton, A., Oh, D., 2011. Modelling dependence in high dimensions with factor copulas. Working paper, Duke University.
Rodriguez, J., 2007. Measuring financial contagion: A copula approach. J. Empir. Finance 14, 401–423.
Rosenberg, J., Schuermann, T., 2006. A general approach to integrated risk management with skewed, fat-tailed risks. J. Financ. Econ. 79, 569–614.
Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. Publ. Inst. Stat. Univ. Paris 8, 229–231.

Empirical Exercises

Open the Chapter9Data.xlsx file from the web site.

1. Replicate the threshold correlations in Figures 9.1 and 9.2. Use a grid of thresholds from 0.15 to 0.85 in increments of 0.01.
2. Simulate 10,000 data points from a bivariate normal distribution to replicate the thresholds in Figure 9.3.
3. Estimate a normal copula model on the S&P 500 and 10-year bond return data. Assume that the marginal distributions have RiskMetrics volatility with symmetric t shocks. Estimate the d parameter for each asset first. Assume that the correlation across the two assets is constant.
4. Simulate 10,000 sets of returns from the model in exercise 3. Compute the 1% *VaR* and *ES* from the model.

The answers to these exercises can be found in the Chapter9Results.xlsx file on the companion site.

For more information see the companion site at
http://www.elsevierdirect.com/companions/9780123744487

10 Option Pricing

1 Chapter Overview

The previous chapters have established a framework for constructing the distribution of a portfolio of assets with simple linear payoffs—for example, stocks, bonds, foreign exchange, forwards, futures, and commodities. This chapter is devoted to the pricing of options. An option derives its value from an underlying asset, but its payoff is not a linear function of the underlying asset price, and so the option price is not a linear function of the underlying asset price either. This nonlinearity adds complications to pricing and risk management.

In this chapter we will do the following:

- Provide some basic definitions and derive a no-arbitrage relationship between put and call prices on the same underlying asset.
- Briefly summarize the binomial tree approach to option pricing.
- Establish an option pricing formula under the simplistic assumption that daily returns on the underlying asset follow a normal distribution with constant variance. We will refer to this as the Black-Scholes-Merton (BSM) formula. While the BSM model provides a useful benchmark, it systematically misprices observed options. We therefore consider the following alternatives.
- Extend the normal distribution model by allowing for skewness and kurtosis in returns. We will rely on the Gram-Charlier expansion around the normal distribution to derive an option pricing formula in this case.
- Extend the model by allowing for time-varying variance relying on the GARCH models from Chapter 4. Two GARCH option pricing models are considered: one allows for general variance specifications, but requires Monte Carlo simulation or another numerical technique; the other assumes a specific variance dynamic but provides a closed-form solution for the option price.
- Introduce the ad hoc implied volatility function (IVF) approach to option pricing. The IVF method is not derived from any coherent theory but it works well in practice.

In this chapter, we will mainly focus attention on the pricing of European options, which can only be exercised on the maturity date. American options that can be

Elements of Financial Risk Management. DOI: 10.1016/B978-0-12-374448-7.00010-5

exercised early will only be discussed briefly. The following chapter will describe in detail the risk management techniques available when the portfolio contains options.

There is enough material in this chapter to fill an entire book, so needless to say the discussion will be brief. We will simply provide an overview of different available option pricing models and suggest further readings at the end of the chapter.

2 Basic Definitions

A European call option gives the owner the right but not the obligation (that is, it gives the *option*) to *buy* a unit of the underlying asset $\tilde{T}$ days from now at the price X. We refer to $\tilde{T}$ as the days to maturity and X as the strike price of the option. We denote the price of the European call option today by c, the price of the underlying asset today by S_t, and at maturity of the option by $S_{t+\tilde{T}}$.

A European put option gives the owner of the option the right to *sell* a unit of the underlying asset $\tilde{T}$ days from now at the price X. We denote the price of the European put option today by p. The European option restricts the owner from exercising the option before the maturity date. American options can be exercised any time before the maturity date.

We note that the number of days to maturity, $\tilde{T}$, is counted in calendar days and not in trading days. A standard year of course has 365 calendar days but only around 252 trading days. In previous chapters, we have been using trading days for returns and Value-at-Risk (*VaR*) horizons, for example, referring to a two-week *VaR* as a 10-day *VaR*. In this chapter it is therefore important to note that we are using 365 days per year when calculating volatilities and interest rates.

The payoff function is the option's defining characteristic. Figure 10.1 contains four panels. The top-left panel shows the payoff from a call option and the top-right panel shows the payoff of a put option both with a strike price of 1137. The payoffs are drawn as a function of the hypothetical price of the underlying asset at maturity of the option, $S_{t+\tilde{T}}$. Mathematically, the payoff function for a call option is

$$Max\left\{S_{t+\tilde{T}} - X, 0\right\}$$

and for a put option it is

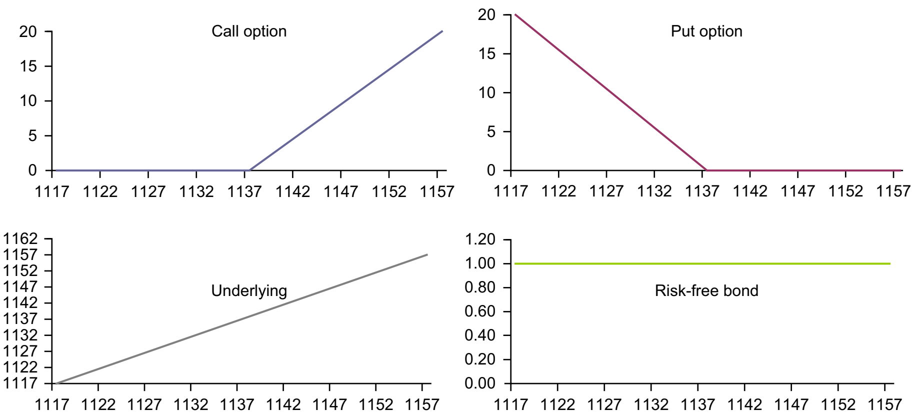
$$Max\left\{X - S_{t+\tilde{T}}, 0\right\}$$

The bottom-left panel of Figure 10.1 shows the payoff function of the underlying asset itself, which is simply a straight line with a slope of one. The bottom right-hand panel shows the value at maturity of a risk-free bond, which pays the face value 1, at maturity $t+\tilde{T}$ regardless of the future price of the underlying risky asset and indeed regardless of any other assets. Notice the linear payoffs of stocks and bonds and the nonlinear payoffs of options.

We next consider the relationship between European call and put option prices. Put-call parity does not rely on any particular option pricing model. It states

$$S_t + p = c + X\exp(-r_f\tilde{T})$$

Figure 10.1 Payoff as a function of the value of the underlying asset at maturity: Call option, put option, underlying asset, and risk-free bond.



Notes: All panels have the future value of the underlying asset on the horizontal axis. The top-left panel plots the call option value, the top right plots the put option value, the bottom left plots the underlying asset itself, and the bottom right plots the risk-free bond.

It can be derived from considering two portfolios: One consists of the underlying asset and the put option and another consists of the call option, and a cash position equal to the discounted value of the strike price. Whether the underlying asset price at maturity, $S_{t+\tilde{T}}$, ends up below or above the strike price X, both portfolios will have the same value, namely $Max\left\{S_{t+\tilde{T}}, X\right\}$, at maturity and therefore they must have the same value today, otherwise arbitrage opportunities would exist: Investors would buy the cheaper of the two portfolios, sell the expensive portfolio, and make risk-free profits. The portfolio values underlying this argument are shown in the following:

Time t	**Time $t+\tilde{T}$**	
Portfolio I	If $S_{t+\tilde{T}} \leq X$	If $S_{t+\tilde{T}} > X$
S_t	$S_{t+\tilde{T}}$	$S_{t+\tilde{T}}$
p	$X - S_{t+\tilde{T}}$	0
$S_t + p$	X	$S_{t+\tilde{T}}$
Portfolio II	If $S_{t+\tilde{T}} \leq X$	If $S_{t+\tilde{T}} > X$
c	0	$S_{t+\tilde{T}} - X$
$X \exp(-r_f \tilde{T})$	X	X
$c + X \exp(-r_f \tilde{T})$	X	$S_{t+\tilde{T}}$

The put-call parity also suggests how options can be used in risk management. Suppose an investor who has an investment horizon of $\tilde{T}$ days owns a stock with current

value S_t. The value of the stock at the maturity of the option is $S_{t+\tilde{T}}$, which in the worst case could be zero. But an investor who owns the stock along with a put option with a strike price of X is guaranteed the future portfolio value $Max\left\{S_{t+\tilde{T}}, X\right\}$, which is at least X. The downside of the stock portfolio including this so-called protective put is thus limited, whereas the upside is still unlimited. The protection is not free however as buying the put option requires paying the current put option price or premium, p.

3 Option Pricing Using Binomial Trees

The key challenge we face when wanting to find a fair value of an option is that it depends on the distribution of the future price of the underlying risky asset (the stock). We begin by making the simplest possible assumption about this distribution, namely that it is binomial. This means that in a short interval of time, the stock price can only take on one of two values, which we can think of as up and down. Clearly this is the simplest possible assumption we can make: If the stock could only take on one possible value going forward then it would not be risky at all. While simple, the binomial tree approach is able to compute the fair market value of American options, which are complicated because early exercise is possible.

The binomial tree option pricing method will be illustrated using the following example: We want to find the fair value of a call and a put option with three months to maturity and a strike price of \$900. The current price of the underlying stock is \$1,000 and the volatility of the log return on the stock is 0.60 or 60% per year corresponding to $0.60/\sqrt{365} = 3.1405\%$ per calendar day.

3.1 Step 1: Build the Tree for the Stock Price

We first must model the distribution of the stock price. The binomial model assumes that the stock price can only take on one of two values at the end of each period. This simple assumption enables us to map out exactly all the possible future values of the stock price. In our example we will assume that the tree has two steps during the six-month maturity of the option, but in practice, a hundred or so steps will be used. The more steps we use, the more accurate the model price will be, but of course the computational burden will increase as well.

Table 10.1 shows how the tree is built in Excel. We know that today's stock price is \$1,000 and so we know the starting point of the tree. We also know the volatility of the underlying stock return (60% per year) and so we know the magnitude of a typical move in the stock price. We need to make sure that the tree accurately reflects the 60% stock return volatility per year.

If the option has three months to maturity and we are building a tree with two steps then each step in the tree corresponds to 1.5 months. The magnitude of the up and down move in each step should therefore reflect a volatility of $0.6\sqrt{dt} = 0.6\sqrt{(3/12)/2} \approx 21.21\%$. In this equation dt denotes the length (in years) of a step in the tree. If we had measured volatility in days then dt should be measured in days as well.

Table 10.1 Building the binomial tree forward from the current stock price

Market Variable				D
$S_t =$	1000			1528.47
Annual $r_f =$	0.05			
Contract Terms				
$X =$	900		B	
$T =$	0.25		1236.31	
Parameters				
Annual Vol =	0.6			
tree steps =	2	A		E
$dt =$	0.125	1000.00		1000.00
$u =$	1.23631111			
$d =$	0.808857893			
			C	
			808.86	
				F
				654.25

Notes: We construct a two-step binomial tree from today's price of $1,000 using an annual volatility of 60%. The total maturity of the tree is three months.

Because we are using log returns a one standard deviation up move corresponds to a gross return of

$$u = \exp(0.2121) = 1.2363$$

and a one standard deviation down move corresponds to a gross return of

$$d = 1/u = \exp(-0.2121) = 0.8089$$

Using these up and down factors the tree is built from the current price of $1,000 on the left side to three potential values in three months, namely $1,528.47 if the stock price moves up twice, $1,000 if it has one up and one down move, and $654.25 if it moves down twice.

3.2 Step 2: Compute the Option Payoff at Maturity

Once we have constructed the tree for the stock price we have three hypothetical stock price values at maturity and we can easily compute the hypothetical call option at each one. The value of an option at maturity is just the payoff stated in the option contract. For a call option we have the payoff function from before:

$$Max\left\{S_{t+\tilde{T}} - X, 0\right\}$$

and so for the three terminal points in the tree in Table 10.1, we get

$$D : Call_D = Max\{1{,}528.47 - 900, 0\} = 428.47$$
$$E : Call_E = Max\{1{,}000.00 - 900, 0\} = 100.00$$
$$F : Call_F = Max\{654.25 - 900, 0\} = 0$$

For the put option we have in general the payoff function

$$Max\left\{X - S_{t+\tilde{T}}, 0\right\}$$

and so in this case we get

$$D : Put_D = Max\{X - S_D, 0\} = Max\{900 - 1{,}528.47, 0\} = 0$$
$$E : Put_E = Max\{X - S_E, 0\} = Max\{900 - 1{,}000, 0\} = 0$$
$$F : Put_F = Max\{X - S_F, 0\} = Max\{900 - 654.25, 0\} = 245.75$$

Table 10.2 shows the three terminal values of the call and put option in the right side of the tree.

The call option values are shown in green font and the put option values are shown in red font.

Table 10.2 Computing the hypothetical option payoffs at maturity

Market Variables				D
$S_t =$	1000			1528.47
Annual $r_f =$	0.05			628.47
				0.00
Contract Terms				
$X =$	900		B	
$T =$	0.25		1236.31	
Parameters				
Annual Vol =	0.6			
tree steps =	2	A		E
$dt =$	0.125	1000.00		1000.00
$u =$	1.23631111			100.00
$d =$	0.808857893			0.00
			C	
Stock is black			808.86	
Call is green				
Put is red				
				F
				654.25
				0.00
				245.75

Notes: For each of the three possible final values of the underlying stock (points D, E, and F) we compute the option value at maturity of the call and put options.

3.3 Step 3: Work Backward in the Tree to Get the Current Option Value

In the tree we have two possible stock price values 1.5 months from now: \$1,236.31 at B and \$808.86 at C. The challenge now is to compute a fair value of the option corresponding to these two stock prices. Consider first point B. We know that going forward from B the stock can only move to either D or E. We know the stock prices at these two points. We also know the option prices at D and E. We need one more piece of information, namely the return on a risk-free bond with 1.5 months to maturity, which corresponds to the length of a step in the tree. The term structure of government debt can be used to obtain this information. Let us assume that the term structure of interest rates is flat at 5% per year.

The key insight is that in a binomial tree we are able to construct a risk-free portfolio using the stock and the option. Because it is risk-free such a portfolio must earn exactly the risk-free rate, which is 5% per year in our example. Consider a portfolio of -1 call option and Δ_B shares of the stock. This means that we have sold one call option and we have bought Δ_B shares of the stock. We need to find a Δ_B such that the portfolio of the option and the stock is risk-free. A portfolio is risk-free if it pays exactly the same in any future state of the world. In our simple binomial world there are only two future states at the end of each step: up and down. Constructing a risk-free portfolio is therefore incredibly simple. Starting from point B we need to find a Δ_B so that

$$S_D \cdot \Delta_B - Call_D = S_E \cdot \Delta_B - Call_E$$

which in this case gives

$$1528.47 \cdot \Delta_B - 628.47 = 1000 \cdot \Delta_B - 100$$

which implies that

$$\Delta_B = \frac{Call_D - Call_E}{S_D - S_E} = \frac{628.47 - 100}{528.47} = 1$$

This shows that we must hold one stock along with the short position of one option in order for the portfolio to be risk-free. The value of this portfolio at D (or E) is \$900 and the portfolio value at B is the discounted value using the risk-free rate for 1.5 months, which is

$$900 \cdot \exp(-0.05 \cdot (3/12)/2) = \$894.39$$

The stock is worth \$1,236.31 at B and so the option must be worth

$$Call_B = 1{,}236.31 - 894.39 = \$341.92$$

which corresponds to the value in green at point B in Table 10.3.

Table 10.3 Working backwards in the tree

Market Variables				
$S_t =$	1000			D
Annual $r_f =$	0.05			1528.47
				628.47
Contract Terms				0.00
$X =$	900			
$T =$	0.25		B	
			1236.31	
Parameters			341.92	
Annual Vol =	0.6		0.00	
tree steps =	2			
$dt =$	0.125	A		E
$u =$	1.23631111	1000.00		1000.00
$d =$	0.808857893	181.47		100.00
RNP =	0.461832245	70.29		0.00
Stock is black			C	
Call is green			808.86	
Put is red			45.90	
			131.43	
				F
				654.25
				0.00
				245.75

Notes: We compute the call and put option values at points B, C, and A using the no-arbitrage principle.

At point C we have instead that

$$1000 \cdot \Delta_C - 100 = 654.25 \cdot \Delta_C - 0$$

so that

$$\Delta_C = \frac{100 - 0}{345.75} = 0.2892$$

This means we have to hold approximately 0.3 shares for each call option we sell. This in turn gives a portfolio value at E (or F) of $1000 \cdot 0.2892 - 100 = \$189.20$. The present value of this is

$$189.20 \cdot \exp(-0.05 \cdot (3/12)/2) = \$188.02$$

At point C we therefore have the call option value

$$Call_C = 0.2892 \cdot 808.86 - 188.02 = \$45.90$$

which is also found in green at point C in Table 10.3.

Now that we have the option prices at points B and C we can construct a risk-free portfolio again to get the option price at point A. We get

$$1236.31 \cdot \Delta_A - 341.92 = 808.86 \cdot \Delta_A - 45.90$$

which implies that

$$\Delta_A = \frac{341.92 - 45.90}{1236.31 - 808.86} = 0.6925$$

which gives a portfolio value at B (or C) of $808.86 \cdot 0.6925 - 45.90 = \$514.24$ with a present value of

$$514.24 \cdot \exp(-0.05 \cdot (3/12)/2) = \$511.04$$

which in turn gives the binomial call option value of

$$c_{Bin} = Call_A = 0.6925 \cdot 1000 - 511.04 = \$181.47$$

which matches the value in Table 10.3. The same computations can be done for a put option. The values are provided in red font in Table 10.3. Once the European call option value has been computed, the put option values can also simply be computed using the put-call parity provided earlier.

3.4 Risk Neutral Valuation

Earlier we priced options based on no-arbitrage arguments: We have constructed a risk-free portfolio that in the absence of arbitrage must earn exactly the risk-free rate. From this portfolio we can back out European option prices. For example, for a call option at point B we used the formula

$$\Delta_B = \frac{Call_D - Call_E}{S_D - S_E} = \frac{Call_D - Call_E}{S_B u - S_B d}$$

which we used to find the call option price at point B using the relationship

$$Call_B = S_B \Delta_B - (S_B u \Delta_B - Call_D) \exp\left(-r_f \cdot dt\right)$$

Using the Δ_B formula we can rewrite the $Call_B$ formula as

$$Call_B = [RNP \cdot Call_D + (1 - RNP) \cdot Call_E] \exp\left(-r_f \cdot dt\right)$$

where the so-called risk neutral probability of an up move is defined as

$$RNP = \frac{\exp\left(r_f \cdot dt\right) - d}{u - d}$$

where dt is defined as before. RNP can be viewed as a probability because the term inside the [*] in the $Call_B$ formula has the form of an expectation of a binomial variable. RNP is termed a risk-neutral probability because the $Call_B$ price appears as a discounted expected value when using RNP in the expectation. Only risk-neutral investors would discount using the risk-free rate and so RNP can be viewed as the probability of an up move in a world where investors are risk neutral.

In our example $dt = (3/12)/2 = 0.125$, $u = 1.2363$, and $d = 0.8089$, so that

$$RNP = \frac{\exp(0.05 \cdot 0.125) - 0.8089}{1.2363 - 0.8089} = 0.4618$$

We can use this number to check that the new formula works. We get

$$\begin{aligned} Call_B &= [RNP \cdot Call_D + (1 - RNP) \cdot Call_E] \exp\left(-r_f \cdot dt\right) \\ &= [0.4618 \cdot 628.47 + (1 - 0.4618) \cdot 100.00] \exp(-0.05 \cdot 0.125) \\ &= 341.92 \end{aligned}$$

just as when using the no-arbitrage argument.

The new formula can be used at any point in the tree. For example at point A we have

$$c_{Bin} = Call_A = [RNP \cdot Call_B + (1 - RNP) \cdot Call_C] \exp\left(-r_f \cdot dt\right)$$

It can also be used for European puts. We have for a put at point C

$$Put_C = [RNP \cdot Put_E + (1 - RNP) \cdot Put_F] \exp\left(-r_f \cdot dt\right)$$

Notice that we again have to work from right to left in the tree when using these formulas. Note also that whereas Δ changes values throughout the tree, RNP is constant throughout the tree.

3.5 Pricing an American Option Using the Binomial Tree

American-style options can be exercised prior to maturity. This added flexibility gives them potentially higher fair market values than European-style options. Fortunately, binomial trees can be used to price American-style options also. We only have to add one calculation in the tree: At the maturity of the option American- and European-style options are equivalent. But at each intermediate point in the tree we must compare the European option value (also known as the continuation value) with the early exercise value and put the largest of the two into the tree at that point.

Consider Table 10.4 where we are pricing an American option that has a strike price of 1,100 but otherwise is exactly the same as the European option considered in Tables 10.1 through 10.3.

Table 10.4 American options: check each node for early exercise

Market Variables				
$S_t =$	1000			D
Annual $r_f =$	0.05			1528.47
				428.47
Contract Terms				0.00
$X =$	1100			
$T =$	0.25		B	
			1236.31	
Parameters			196.65	
Annual Vol =	0.6		53.48	
tree steps =	2			
$dt =$	0.125	A		E
$u =$	1.23631111	1000.00		1000.00
$d =$	0.808857893	90.25		0.00
RNP =	0.461832245	180.25		100.00
Stock is black			C	
American call is green			808.86	
American put is red			0.00	
			291.14	
				F
				654.25
				0.00
				445.75

Notes: We compute the American option values by checking for early exercise at each point in the tree.

If we exercise the American put option at point C we get

$$Max\{1,100 - 808.86, 0\} = \$291.14$$

Let us now compute the European put value at this point. Using the previous method we have the risk-neutral probability of an up-move $RNP = 0.4618$, so that the European put value at point C is

$$\begin{aligned} Put_C &= [RNP \cdot Put_E + (1 - RNP) \cdot Put_F] \exp\left(-r_f \cdot dt\right) \\ &= \$284.29 \end{aligned}$$

which is of course lower than the early exercise value \$284.29. Early exercise of the put is optimal at point C and the fair market value of the American option is therefore \$291.14 at this point. This value will now influence the American put option value at point A, which will also be larger than its corresponding European put option value. Table 10.4 shows that the American put is worth \$180.25 at point A.

The American call option price is \$90.25, which turns out to be the European call option price as well. This is because American call stock options should only be

exercised early if a large cash dividend is imminent. In our example there were no dividends and so early exercise of the American call is never optimal, which in turn makes the American call option price equal to the European call option price.

3.6 Dividend Flows, Foreign Exchange, and Futures Options

In the case where the underlying asset pays out a stream of dividends or other cash flows we need to adjust the *RNP* formula. Consider an underlying stock index that pays out cash at a rate of q per year. In this case we have

$$RNP = \frac{\exp\left(\left(r_f - q\right) \cdot dt\right) - d}{u - d}$$

When the underlying asset is a foreign exchange rate then q is set to the interest rate of the foreign currency. When the underlying asset is a futures contract then $q = r_f$ so that $RNP = (1 - d) / (u - d)$ for futures options.

4 Option Pricing under the Normal Distribution

The binomial tree approach is very useful because it is so simple to derive and because it allows us to price American as well as European options. A downside of binomial tree pricing is that we do not obtain a closed-form formula for the option price.

In order to do so we now assume that daily returns on an asset be independently and identically distributed according to the normal distribution,

$$R_{t+1} = \ln(S_{t+1}) - \ln(S_t) \overset{i.i.d.}{\sim} N\left(\mu - \frac{1}{2}\sigma^2, \sigma^2\right)$$

Then the aggregate return over $\tilde{T}$ days will also be normally distributed with the mean and variance appropriately scaled as in

$$R_{t+1:t+\tilde{T}} = \ln\left(S_{t+\tilde{T}}\right) - \ln(S_t) \sim N\left(\tilde{T}\left(\mu - \frac{1}{2}\sigma^2\right), \tilde{T}\sigma^2\right)$$

and the future asset price can of course be written as

$$S_{t+\tilde{T}} = S_t \exp\left(R_{t+1:t+\tilde{T}}\right)$$

The risk-neutral valuation principle calculates the option price as the discounted expected payoff, where discounting is done using the risk-free rate and where the expectation is taken using the risk-neutral distribution:

$$c = \exp(-r_f \tilde{T}) E_t^* \left[Max\left\{S_{t+\tilde{T}} - X, 0\right\}\right]$$

where $Max\left\{S_{t+\tilde{T}} - X, 0\right\}$ as before is the payoff function and where r_f is the risk-free interest rate per day. The expectation $E_t^*[*]$ is taken using the risk-neutral distribution

where all assets earn an expected return equal to the risk-free rate. In this case the option price can be written as

$$
\begin{aligned}
c &= \exp(-r_f\tilde{T}) \int_{-\infty}^{\infty} Max\left\{S_t \exp(x^*) - X, 0\right\} f(x^*)dx^* \\
&= \exp(-r_f\tilde{T}) \int_{\ln(X/S_t)}^{\infty} S_t \exp(x^*) f(x^*)dx^* - \int_{\ln(X/S_t)}^{\infty} Xf(x^*)dx^*
\end{aligned}
$$

where x^* is the risk-neutral variable corresponding to the underlying asset return between now and the maturity of the option. $f(x^*)$ denotes the risk-neutral distribution, which we take to be the normal distribution so that $x^* \sim N(\tilde{T}(r_f - \frac{1}{2}\sigma^2), \tilde{T}\sigma^2)$. The second integral is easily evaluated whereas the first requires several steps. In the end we obtain the Black-Scholes-Merton (BSM) call option price

$$
\begin{aligned}
c_{BSM} &= \exp(-r_f\tilde{T}) \left[S_t \exp(r_f\tilde{T})\Phi(d) - X\Phi\left(d - \sigma\sqrt{\tilde{T}}\right)\right] \\
&= S_t\Phi(d) - \exp(-r_f\tilde{T})X\Phi\left(d - \sigma\sqrt{\tilde{T}}\right)
\end{aligned}
$$

where $\Phi(\bullet)$ is the cumulative density of a standard normal variable, and where

$$
d = \frac{\ln(S_t/X) + \tilde{T}\left(r_f + \sigma^2/2\right)}{\sigma\sqrt{\tilde{T}}}
$$

Black, Scholes, and Merton derived this pricing formula in the early 1970s using a model where trading takes place in continuous time when assuming continuous trading only the absence of arbitrage opportunities is needed to derive the formula.

It is worth emphasizing that to stay consistent with the rest of the book, the volatility and risk-free interest rates are both denoted in daily terms, and option maturity is denoted in number of calendar days, as this is market convention.

The elements in the option pricing formula have the following interpretation:

- $\Phi(d - \sigma\sqrt{\tilde{T}})$ is the risk-neutral probability of exercise.
- $X\Phi(d - \sigma\sqrt{\tilde{T}})$ is the expected risk-neutral payout when exercising.
- $S_t\Phi(d)\exp(r_f\tilde{T})$ is the risk-neutral expected value of the stock acquired through exercise of the option.
- $\Phi(d)$ measures the sensitivity of the option price to changes in the underlying asset price, S_t, and is referred to as the delta of the option, where $\delta_{BSM} \equiv \frac{\partial c_{BSM}}{\partial S_t}$ is the first derivative of the option with respect to the underlying asset price. This and other sensitivity measures are discussed in detail in the next chapter.

Using the put-call parity result and the formula for c_{BSM}, we can get the put price formula as

$$\begin{aligned}p_{BSM} &= c_{BSM} + X\exp(-r_f\tilde{T}) - S_t \\ &= e^{-r_f\tilde{T}}\left\{X\left[1-\Phi\left(d-\sigma\sqrt{\tilde{T}}\right)\right] - S_t\left[1-\Phi(d)\right]e^{r_f\tilde{T}}\right\} \\ &= e^{-r_f\tilde{T}}X\Phi\left(\sigma\sqrt{\tilde{T}}-d\right) - S_t\Phi(-d)\end{aligned}$$

where the last line comes from the symmetry of the normal distribution, which implies that $[1-\Phi(z)] = \Phi(-z)$ for any value of z.

In the case where cash flows such as dividends accrue to the underlying asset, we discount the current asset price to account for the cash flows by replacing S_t by $S_t\exp(-q\tilde{T})$ everywhere, where q is the expected rate of cash flow per day until maturity of the option. This adjustment can be made to both the call and the put price formula, and in both cases the formula for d will then be

$$d = \frac{\ln(S_t/X) + \tilde{T}\left(r_f - q + \sigma^2/2\right)}{\sigma\sqrt{\tilde{T}}}$$

The adjustment is made because the option holder at maturity receives only the underlying asset on that date and not the cash flow that has accrued to the asset during the life of the option. This cash flow is retained by the owner of the underlying asset.

We now want to use the Black-Scholes pricing model to price a European call option written on the S&P 500 index. On January 6, 2010, the value of the index was 1137.14. The European call option has a strike price of 1110 and 43 days to maturity. The risk-free interest rate for a 43-day holding period is found from the T-bill rates to be 0.0006824% per day (that is, 0.000006824) and the dividend accruing to the index over the next 43 days is expected to be 0.0056967% per day. For now, we assume the volatility of the index is 0.979940% per day. Thus, we have

$$\begin{aligned}S_t &= 1137.14 \\ X &= 1110 \\ \tilde{T} &= 43 \\ r_f &= 0.0006824\% \\ q &= 0.0056967\% \\ \sigma &= 0.979940\%\end{aligned}$$

and we can calculate

$$d = \frac{\ln(S_t/X) + \tilde{T}\left(r_f - q + \sigma^2/2\right)}{\sigma\sqrt{\tilde{T}}} = 0.374497, \text{ and } d - \sigma\sqrt{\tilde{T}} = 0.310238$$

which gives

$$\Phi(d) = 0.645983, \text{ and } \Phi\left(d-\sigma\sqrt{\tilde{T}}\right) = 0.621810$$

from which we can calculate the BSM call option price as

$$c_{BSM} = S_t \exp(-q\tilde{T})\Phi(d) - \exp(-r_f\tilde{T})X\Phi\left(d - \sigma\sqrt{\tilde{T}}\right) = 42.77$$

4.1 Model Implementation

The simple BSM model implies that a European option price can be written as a nonlinear function of six variables,

$$c_{BSM} = c(S_t, r_f, X, \tilde{T}, q; \sigma)$$

The stock price is readily available, and a treasury bill rate with maturity $\tilde{T}$ can be used as the risk-free interest rate. The strike price and time to maturity are known features of any given option contract, thus only one parameter needs to be estimated—namely, the volatility, σ. As the option pricing formula is nonlinear, volatility can be estimated from a sample of n options on the same underlying asset, minimizing the mean-squared dollar pricing error (MSE):

$$MSE_{BSM} = \min_{\sigma}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(c_i^{mkt} - c_{BSM}(S_t, r_f, X_i, \tilde{T}_i, q; \sigma)\right)^2\right\}$$

where c_i^{mkt} denotes the observed market price of option i. The web site that contains answers to the exercises at the end of this chapter includes an example of this numerical optimization. Notice that we also could, of course, simply have plugged in an estimate of σ from returns on the underlying asset; however, using the observed market prices of options tends to produce much more accurate model prices.

Using prices on a sample of 103 call options traded on the S&P 500 index on January 6, 2010, we estimate the volatility, which minimizes the MSE to be 0.979940% per day. This was the volatility estimate used in the numerical pricing example. Further details of this calculation can be found on the web page.

4.2 Implied Volatility

From Chapter 1, we know that the assumption of daily asset returns following the normal distribution is grossly violated in the data. We therefore should worry that an option pricing theory based on the normal distribution will not offer an appropriate description of reality. To assess the quality of the normality-based model, consider the so-called implied volatility calculated as

$$\sigma_{BSM}^{iv} = c_{BSM}^{-1}\left(S_t, r_f, X, \tilde{T}, q, c^{mkt}\right)$$

where c^{mkt} again denotes the observed market price of the option, and where $c_{BSM}^{-1}(*)$ denotes the inverse of the BSM option pricing formula derived earlier. The implied volatilities can be found contract by contract by using a numerical equation solver.

Returning to the preceding numerical example of the S&P 500 call option traded on January 6, 2010, knowing that the actual market price for the option was 42.53, we can calculate the implied volatility to be

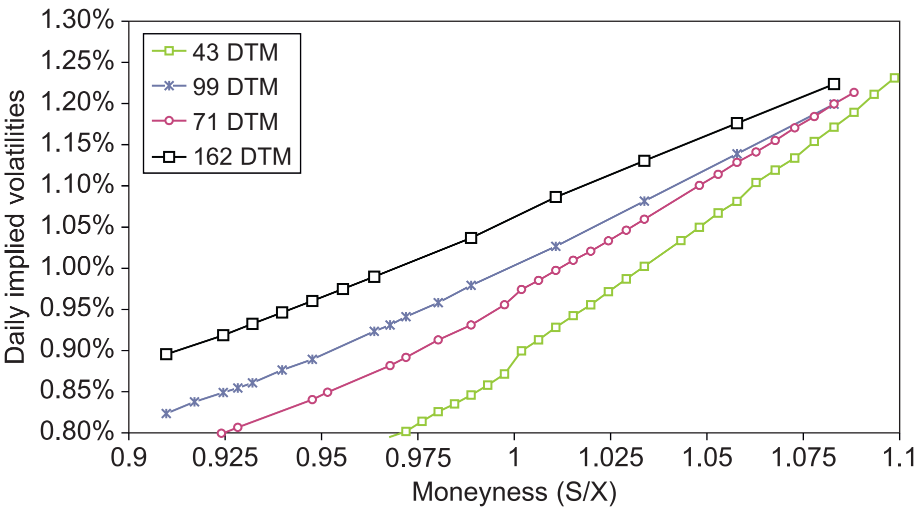
$$\sigma^{iv}_{BSM} = c^{-1}_{BSM}\left(S_t, r_f, X, \tilde{T}, q, 42.53\right) = 0.971427\%$$

where the S_t, r_f, X, $\tilde{T}$, and q variables are as in the preceding example. The 0.971427% volatility estimate is such that if we had used it in the BSM formula, then the model price would have equalled the market price exactly; that is,

$$42.53 = c_{BSM}\left(S_t, r_f, X, \tilde{T}, q, 0.971427\%\right)$$

If the normality assumption imposed on the model were true, then the implied volatility should be roughly constant across strike prices and maturities. However, actual option data displays systematic patterns in implied volatility, thus violating the normality-based option pricing theory. Figure 10.2 shows the implied volatility of various S&P 500 index call options plotted as a function of moneyness (S/X) on January 6, 2010. The picture shows clear evidence of the so-called *smirk*. Furthermore, the smirk is most evident at shorter horizons. As we will see shortly, this smirk can arise from skewness in the underlying distribution, which is ignored in the BSM model relying on normality. Options on foreign exchange tend to show a more symmetric pattern of implied volatility, which is referred to as the *smile*. The smile can arise from kurtosis in the underlying distribution, which is again ignored in the BSM model.

Figure 10.2 Implied BSM daily volatility from S&P 500 index options with 43, 99, 71, and 162 days to maturity (DTM) quoted on January 06, 2010.



Notes: We plot one day's BSM implied volatilities against moneyness. Each line corresponds to a specific maturity.

Smirk and smile patterns in implied volatility constitute evidence of misspecification in the BSM model. Consider for example pricing options with the BSM formula using a daily volatility of approximately 1% for all options. In Figure 10.2, the implied volatility is approximately 1% for at-the-money options for which $S/X \approx 1$. Therefore, the BSM price would be roughly correct for these options. However, for options that are in-the-money—that is, $S/X > 1$—the BSM implied volatility is higher than 1%, which says that the BSM model needs a higher than 1% volatility to fit the market data. This is because option prices are increasing in the underlying volatility. Using the BSM formula with a volatility of 1% would result in a BSM price that is too low. The BSM is thus said to underprice in-the-money call options. From the put-call parity formula, we can conclude that the BSM model also underprices out-of-the-money put options.

5 Allowing for Skewness and Kurtosis

We now introduce a relatively simple model that is capable of making up for some of the obvious mispricing in the BSM model. We again have one day returns defined as

$$R_{t+1} = \ln(S_{t+1}) - \ln(S_t)$$

and $\tilde{T}$-period returns as

$$R_{t+1:t+\tilde{T}} = \ln(S_{t+\tilde{T}}) - \ln(S_t)$$

The mean and variance of the daily returns are again defined as $E(R_{t+1}) = \mu - \frac{1}{2}\sigma^2$ and $E\left(R_{t+1} - \mu + \frac{1}{2}\sigma^2\right)^2 = \sigma^2$. We previously defined skewness by ζ_1. We now explicitly define skewness of the one-day return as

$$\zeta_{11} = \frac{E\left(R_{t+1} - \mu + \frac{1}{2}\sigma^2\right)^3}{\sigma^3}$$

Skewness is informative about the degree of asymmetry of the distribution. A negative skewness arises from large negative returns being observed more frequently than large positive returns. Negative skewness is a stylized fact of equity index returns, as we saw in Chapter 1. Kurtosis of the one-day return is now defined as

$$\zeta_{21} = \frac{E\left(R_{t+1} - \mu + \frac{1}{2}\sigma^2\right)^4}{\sigma^4} - 3$$

which is sometimes referred to as excess kurtosis due to the subtraction by 3. Kurtosis tells us about the degree of tail fatness in the distribution of returns. If large (positive or negative) returns are more likely to occur in the data than in the normal distribution, then the kurtosis is positive. Asset returns typically have positive kurtosis.

Assuming that returns are independent over time, the skewness at horizon $\tilde{T}$ can be written as a simple function of the daily skewness,

$$\zeta_{1\tilde{T}} = \zeta_{11}/\sqrt{\tilde{T}}$$

and correspondingly for kurtosis

$$\zeta_{2\tilde{T}} = \zeta_{21}/\tilde{T}$$

Notice that both skewness and kurtosis will converge to zero as the return horizon, $\tilde{T}$, and thus the maturity of the option increases. This corresponds well with the implied volatility in Figure 10.2, which displayed a more pronounced smirk pattern for short-term as opposed to long-term options.

We now define the standardized return at the $\tilde{T}$-day horizon as

$$w_{\tilde{T}} = \frac{R_{t+1:t+\tilde{T}} - \tilde{T}\left(\mu - \frac{1}{2}\sigma^2\right)}{\sqrt{\tilde{T}}\sigma}$$

so that

$$R_{t+1:t+\tilde{T}} = \left(\mu - \frac{1}{2}\sigma^2\right)\tilde{T} + \sigma\sqrt{\tilde{T}}w_{\tilde{T}}$$

and assume that the standardized returns follow the distribution given by the Gram-Charlier expansion, which is written as

$$f\left(w_{\tilde{T}}\right) = \phi\left(w_{\tilde{T}}\right) - \zeta_{1\tilde{T}}\frac{1}{3!}D^3\phi\left(w_{\tilde{T}}\right) + \zeta_{2\tilde{T}}\frac{1}{4!}D^4\phi\left(w_{\tilde{T}}\right)$$

where $3! = 3 \cdot 2 \cdot 1 = 6$, $\phi\left(w_{\tilde{T}}\right)$ is the standard normal density, and D^j is its jth derivative. We have

$$\begin{aligned}
D^1\phi(z) &= -z\phi(z) \\
D^2\phi(z) &= (z^2 - 1)\phi(z) \\
D^3\phi(z) &= -(z^3 - 3z)\phi(z) \\
D^4\phi(z) &= (z^4 - 6z^2 + 3)\phi(z)
\end{aligned}$$

The Gram-Charlier density function $f\left(w_{\tilde{T}}\right)$ is an expansion around the normal density function, $\phi(w_{\tilde{T}})$, allowing for a nonzero skewness, $\zeta_{1\tilde{T}}$, and kurtosis $\zeta_{2\tilde{T}}$. The Gram-Charlier expansion can approximate a wide range of densities with nonzero higher moments, and it collapses to the standard normal density when skewness and kurtosis are both zero. We notice the similarities with the Cornish-Fisher expansion for Value-at-Risk in Chapter 6, which is a similar expansion, but for the inverse cumulative density function instead of the density function itself.

To price European options, we can again write the generic risk-neutral call pricing formula as

$$c = e^{-r_f \tilde{T}} E_t^* \left[Max \left\{ S_{t+\tilde{T}} - X, 0 \right\} \right]$$

Thus, we must solve

$$c = e^{-r_f \tilde{T}} \int_{\ln X/S_t}^{\infty} \left(S_t \exp(x^*) - X \right) f\left(x^*\right) dx^*$$

Earlier we relied on x^* following the normal distribution with mean $r_f - \frac{1}{2}\sigma^2$ and variance σ^2 per day. But we now instead define the standardized risk-neutral return at horizon $\tilde{T}$ as

$$w_{\tilde{T}}^* = \frac{\left(x^* - \left(r_f - \frac{1}{2}\sigma^2 \right) \tilde{T} \right)}{\sqrt{\tilde{T}}\sigma}$$

and assume it follows the Gram-Charlier (GC) distribution.

In this case, the call option price can be derived as being approximately equal to

$$\begin{aligned} c_{GC} &\approx S_t \Phi(d) - X e^{-r_f \tilde{T}} \Phi\left(d - \sqrt{\tilde{T}}\sigma \right) \\ &\quad + S_t \phi(d) \sqrt{\tilde{T}}\sigma \left[\frac{\zeta_{1\tilde{T}}}{3!} \left(2\sqrt{\tilde{T}}\sigma - d \right) - \frac{\zeta_{2\tilde{T}}}{4!} \left(1 - d^2 + 3d\sqrt{\tilde{T}}\sigma - 3\tilde{T}\sigma^2 \right) \right] \\ &= S_t \Phi(d) - X e^{-r_f \tilde{T}} \Phi\left(d - \sqrt{\tilde{T}}\sigma \right) \\ &\quad + S_t \phi(d) \sigma \left[\frac{\zeta_{11}}{3!} \left(2\sqrt{\tilde{T}}\sigma - d \right) - \frac{\zeta_{21}/\sqrt{\tilde{T}}}{4!} \left(1 - d^2 + 3d\sqrt{\tilde{T}}\sigma - 3\tilde{T}\sigma^2 \right) \right] \end{aligned}$$

where we have substituted in for skewness using $\zeta_{1\tilde{T}} = \zeta_{11}/\sqrt{\tilde{T}}$ and for kurtosis using $\zeta_{2\tilde{T}} = \zeta_{21}/\tilde{T}$. We will refer to this as the GC option pricing model. The approximation comes from setting the terms involving σ^3 and σ^4 to zero, which also enables us to use the definition of d from the BSM model. Using this approximation, the GC model is just the simple BSM model plus additional terms that vanish if there is neither skewness ($\zeta_{11} = 0$) nor kurtosis ($\zeta_{21} = 0$) in the data. The GC formula can be extended to allow for a cash flow q in the same manner as the BSM formula shown earlier.

5.1 Model Implementation

This GC model has three unknown parameters: σ, ζ_{11}, and ζ_{21}. They can be estimated as before using a numerical optimizer minimizing the mean squared error

$$MSE_{GC} = \min_{\sigma, \zeta_{11}, \zeta_{21}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(c_i^{mkt} - c_{GC}(S_t, r_f, X_i, \tilde{T}_i; \sigma, \zeta_{11}, \zeta_{21}) \right)^2 \right\}$$

We can calculate the implied BSM volatilities from the GC model prices by

$$\sigma^{iv}_{GC} = c^{-1}_{BSM}\left(S_t, r_f, X, \tilde{T}, c_{GC}\right)$$

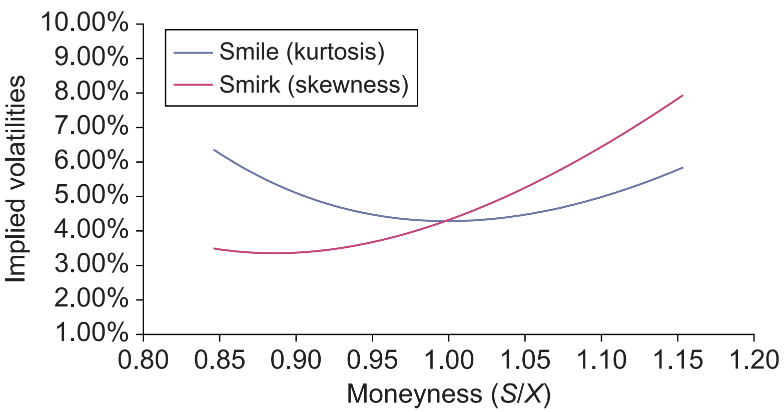
where $c^{-1}_{BSM}(*)$ is the inverse of the BSM model with respect to volatility. But we can also rely on the following approximate formula for daily implied BSM volatility:

$$\sigma^{iv}_{GC} = c^{-1}_{BSM}\left(S_t, r_f, X, \tilde{T}, c_{GC}\right) \approx \sigma\left[1 - \frac{\zeta_{11}/\sqrt{\tilde{T}}}{3!}d - \frac{\zeta_{21}/\tilde{T}}{4!}\left(1 - d^2\right)\right]$$

Notice this is just volatility times an additional term, which equals one if there is no skewness or kurtosis. Figure 10.3 plots two implied volatility curves for options with 10 days to maturity. One has a skewness of −3 and a kurtosis of 7 and shows the smirk, and the other has no skewness but a kurtosis of 8 and shows a smile.

The main advantages of the GC option pricing framework are that it allows for deviations from normality, that it provides closed-form solutions for option prices, and, most important, it is able to capture the systematic patterns in implied volatility found in observed option data. For example, allowing for negative skewness implies that the GC option price will be higher than the BSM price for in-the-money calls, thus removing the tendency for BSM to underprice in-the-money calls, which we saw in Figure 10.2.

Figure 10.3 Implied BSM volatility from Gram-Charlier model prices.



Notes: We plot the implied BSM volatility for options with 10 days to maturity using the Gram-Charlier model. The red line has a skewness of −3 and a kurtosis of 7. The blue line has a skewness of 0 and a kurtosis of 8.

6 Allowing for Dynamic Volatility

While the GC model is capable of capturing implied volatility smiles and smirks at a given point in time, it assumes that volatility is constant over time and is thus inconsistent with the empirical observations we made earlier. Put differently, the GC model is able to capture the strike price structure but not the maturity structure in observed options prices. In Chapters 4 and 5 we saw that variance varies over time in a predictable fashion: High-variance days tend to be followed by high-variance days and vice versa, which we modeled using GARCH and other types of models. When returns are independent, the standard deviation of returns at the $\tilde{T}$-day horizon is simply $\sqrt{\tilde{T}}$ times the daily volatility, whereas the GARCH model implies that the term structure of variance depends on the variance today and does not follow the simple square root rule.

We now consider option pricing allowing for the underlying asset returns to follow a GARCH process. The GARCH option pricing model assumes that the expected return on the underlying asset is equal to the risk-free rate, r_f, plus a premium for volatility risk, λ, as well as a normalization term. The observed daily return is then equal to the expected return plus a noise term. The noise term is conditionally normally distributed with mean zero and variance following a GARCH(1,1) process with leverage as in Chapter 4. By letting the past return feed into variance in a magnitude depending on the sign of the return, the leverage effect creates an asymmetry in the distribution of returns. This asymmetry is important for capturing the skewness implied in observed option prices.

Specifically, we can write the return process as

$$R_{t+1} \equiv \ln(S_{t+1}) - \ln(S_t) = r_f + \lambda\sigma_{t+1} - \frac{1}{2}\sigma^2_{t+1} + \sigma_{t+1}z_{t+1}$$
$$\text{with } z_{t+1} \sim N(0,1), \text{ and } \sigma^2_{t+1} = \omega + \alpha\left(\sigma_t z_t - \theta\sigma_t\right)^2 + \beta\sigma^2_t$$

Notice that the expected value and variance of tomorrow's return conditional on all the information available at time t are

$$E_t\left[R_{t+1}\right] = r_f + \lambda\sigma_{t+1} - \frac{1}{2}\sigma^2_{t+1}$$
$$V_t\left[R_{t+1}\right] = \sigma^2_{t+1}$$

For a generic normally distributed variable $x \sim N(\mu, \sigma^2)$, we have that $E[\exp(x)] = \exp\left(\mu + \sigma^2/2\right)$ and therefore we get

$$\begin{aligned} E_t\left[S_{t+1}/S_t\right] &= E_t\left[\exp\left(r_f + \lambda\sigma_{t+1} - \frac{1}{2}\sigma^2_{t+1} + \sigma_{t+1}z_{t+1}\right)\right] \\ &= \exp\left(r_f + \lambda\sigma_{t+1} - \frac{1}{2}\sigma^2_{t+1}\right)E_t\left[\exp\left(\sigma_{t+1}z_{t+1}\right)\right] \end{aligned}$$

$$= \exp\left(r_f + \lambda\sigma_{t+1} - \frac{1}{2}\sigma_{t+1}^2\right)\exp\left(\frac{1}{2}\sigma_{t+1}^2\right)$$
$$= \exp\left(r_f + \lambda\sigma_{t+1}\right)$$

where we have used $\sigma_{t+1}z_{t+1} \sim N(0, \sigma_{t+1}^2)$. This expected return equation highlights the role of λ as the price of volatility risk.

We can again solve for the option price using the risk-neutral expectation as in

$$c = \exp(-r_f\tilde{T})E_t^*\left[Max\left\{S_{t+\tilde{T}} - X, 0\right\}\right]$$

Under risk neutrality, we must have that

$$E_t^*\left[S_{t+1}/S_t\right] = \exp\left(r_f\right)$$
$$V_t^*\left[R_{t+1}\right] = \sigma_{t+1}^2$$

so that the expected rate of return on the risky asset equals the risk-free rate and the conditional variance under risk neutrality is the same as the one under the original process. Consider the following process:

$$R_{t+1} \equiv \ln(S_{t+1}) - \ln(S_t) = r_f - \frac{1}{2}\sigma_{t+1}^2 + \sigma_{t+1}z_{t+1}^*$$
$$\text{with } z_{t+1}^* \sim N(0,1), \text{ and } \sigma_{t+1}^2 = \omega + \alpha\left(\sigma_t z_t^* - \lambda\sigma_t - \theta\sigma_t\right)^2 + \beta\sigma_t^2$$

In this case, we can check that the conditional mean equals

$$E_t^*\left[S_{t+1}/S_t\right] = E_t^*\left[\exp\left(r_f - \frac{1}{2}\sigma_{t+1}^2 + \sigma_{t+1}z_{t+1}^*\right)\right]$$
$$= \exp\left(r_f - \frac{1}{2}\sigma_{t+1}^2\right)E_t^*\left[\exp\left(\sigma_{t+1}z_{t+1}^*\right)\right]$$
$$= \exp\left(r_f - \frac{1}{2}\sigma_{t+1}^2\right)\exp\left(\frac{1}{2}\sigma_{t+1}^2\right)$$
$$= \exp\left(r_f\right)$$

which satisfies the first condition. Furthermore, the conditional variance under the risk-neutral process equals

$$V_t^*\left[R_{t+1}\right] = E_t^*\left[\omega + \alpha\left(\sigma_t z_t^* - \lambda\sigma_t - \theta\sigma_t\right)^2 + \beta\sigma_t^2\right]$$
$$= E_t\left[\omega + \alpha\left(R_t - r_f + \frac{1}{2}\sigma_{t+1}^2 - \lambda\sigma_t - \theta\sigma_t\right)^2 + \beta\sigma_t^2\right]$$
$$= E_t\left[\omega + \alpha\left(\sigma_t z_t - \theta\sigma_t\right)^2 + \beta\sigma_t^2\right]$$
$$= \sigma_{t+1}^2$$

where the last equality comes from tomorrow's variance being known at the end of today in the GARCH model. The conclusion is that the conditions for a risk-neutral process are met.

An advantage of the GARCH option pricing approach introduced here is its flexibility: The previous analysis could easily be redone for any of the GARCH variance models introduced in Chapter 4. More important, it is able to fit observed option prices quite well.

6.1 Model Implementation

While we have found a way to price the European option under risk neutrality, unfortunately, we do not have a closed-form solution available. Instead, we have to use simulation to calculate the price

$$c = \exp(-r_f \tilde{T}) E_t^* \left[Max\left\{ S_{t+\tilde{T}} - X, 0 \right\} \right]$$

The simulation can be done as follows: First notice that we can get rid of a parameter by writing

$$\begin{aligned} \sigma_{t+1}^2 &= \omega + \alpha \left(\sigma_t z_t^* - \lambda \sigma_t - \theta \sigma_t \right)^2 + \beta \sigma_t^2 \\ &= \omega + \alpha \left(\sigma_t z_t^* - \lambda^* \sigma_t \right)^2 + \beta \sigma_t^2, \quad \text{with } \lambda^* \equiv \lambda + \theta \end{aligned}$$

Now, for a given conditional variance σ_{t+1}^2, and parameters $\omega, \alpha, \beta, \lambda^*$, we can use Monte Carlo simulation as in Chapter 8 to create future hypothetical paths of the asset returns. Parameter estimation will be discussed subsequently. Graphically, we can illustrate the simulation of hypothetical daily returns from day $t+1$ to the maturity on day $t+\tilde{T}$ as

$$\begin{array}{llllll} & \breve{z}_{1,1}^* \to \breve{R}_{1,t+1}^* \to \breve{\sigma}_{1,t+2}^2 & \breve{z}_{1,2}^* \to \breve{R}_{1,t+2}^* \to \breve{\sigma}_{1,t+3}^2 & \cdots & \breve{z}_{1,\tilde{T}}^* \to \breve{R}_{1,t+\tilde{T}}^* \\ \nearrow & \breve{z}_{2,1}^* \to \breve{R}_{2,t+1}^* \to \breve{\sigma}_{2,t+2}^2 & \breve{z}_{2,2}^* \to \breve{R}_{2,t+2}^* \to \breve{\sigma}_{2,t+3}^2 & \cdots & \breve{z}_{2,\tilde{T}}^* \to \breve{R}_{2,t+\tilde{T}}^* \\ \sigma_{t+1}^2 \longrightarrow & \cdots & \cdots & \cdots & \cdots \\ \searrow & \cdots & \cdots & \cdots & \cdots \\ & \breve{z}_{MC,1}^* \to \breve{R}_{MC,t+1}^* \to \breve{\sigma}_{MC,t+2}^2 & \breve{z}_{MC,2}^* \to \breve{R}_{MC,t+2}^* \to \breve{\sigma}_{MC,t+3}^2 & \cdots & \breve{z}_{MC,\tilde{T}}^* \to \breve{R}_{MC,t+\tilde{T}}^* \end{array}$$

where the $\breve{z}_{i,j}^*$s are obtained from a $N(0,1)$ random number generator and where MC is the number of simulated return paths. We need to calculate the expectation term $E_t^*[*]$ in the option pricing formula using the risk-neutral process, thus, we calculate the simulated risk-neutral return in period $t+j$ for simulation path i as

$$\breve{R}_{i,t+j}^* = r_f - \frac{1}{2} \breve{\sigma}_{i,t+j}^2 + \breve{\sigma}_{i,t+j} \breve{z}_{i,j}^*$$

and the variance is updated by

$$\breve{\sigma}_{i,t+j+1}^2 = \omega + \alpha \left(\breve{\sigma}_{i,t+j} \breve{z}_{i,j}^* - \lambda^* \breve{\sigma}_{i,t+j} \right)^2 + \beta \breve{\sigma}_{i,t+j}^2$$

As in Chapter 8, the simulation paths in the first period all start out from the same σ_{t+1}^2; therefore, we have

$$\check{R}_{i,t+1}^* = r_f - \frac{1}{2}\sigma_{t+1}^2 + \sigma_{t+1}\check{z}_{i,1}^*$$
$$\sigma_{i,t+2}^2 = \omega + \alpha\left(\sigma_{t+1}\check{z}_{i,1}^* - \lambda^*\sigma_{t+1}\right)^2 + \beta\sigma_{t+1}^2$$

for all i.

Once we have simulated, say, 5000 paths ($MC = 5000$) each day until the maturity date $\tilde{T}$, we can calculate the hypothetical risk-neutral asset prices at maturity as

$$\check{S}_{i,t+\tilde{T}}^* = S_t \exp\left(\sum_{j=1}^{\tilde{T}} \check{R}_{i,t+j}^*\right), \quad i = 1, 2, \ldots, MC$$

and the option price is calculated taking the average over the future hypothetical payoffs and discounting them to the present as in

$$c_{GH} = \exp(-r_f\tilde{T})E_t^*\left[Max\left\{S_{t+\tilde{T}} - X, 0\right\}\right]$$
$$\approx \exp(-r_f\tilde{T})\frac{1}{MC}\sum_{i=1}^{MC} Max\left\{\check{S}_{i,t+\tilde{T}}^* - X, 0\right\}$$

where GH denotes GARCH.

Thus, we are using simulation to calculate the average future payoff, which is then used as an estimate of the expected value, $E_t^*[*]$. As the number of Monte Carlo replications gets infinitely large, the average will converge to the expectation. In practice, around 5000 replications suffice to get a reasonably precise estimate. The web site accompanying this book contains a spreadsheet with a Monte Carlo simulation calculating GARCH option prices.

In theory, we could, of course, estimate all the parameters in the GARCH model using the maximum likelihood method from Chapter 4 on the underlying asset returns. But to obtain a better fit of the option prices, we can instead minimize the option pricing errors directly. Treating the initial variance σ_{t+1}^2 as a parameter to be estimated, we can estimate the GARCH option pricing model on a daily sample of options by numerically minimizing the mean squared error

$$MSE_{GH} = \min_{\sigma_{t+1}^2, \omega, \alpha, \beta, \lambda^*} \left\{\frac{1}{n}\sum_{i=1}^{n}\left(c_i^{mkt} - c_{GH}(S_t, r_f, X_i, \tilde{T}_i; \sigma_{t+1}^2, \omega, \alpha, \beta, \lambda^*)\right)^2\right\}$$

Alternatively, an objective function based on implied volatility can be used. Notice that for every new parameter vector the numerical optimizer tries, the GARCH options must all be repriced using the MC simulation technique, thus the estimation can be quite time consuming.

6.2 A Closed-Form GARCH Option Pricing Model

A significant drawback of the GARCH option pricing framework outlined here is clearly that it does not provide us with a closed-form solution for the option price, which must instead be calculated through simulation. Although the simulation technique is straightforward, it does take computing time and introduces an additional source of error arising from the approximation of the simulated average to the expected value.

Fortunately, if we are willing to accept a particular type of GARCH process, then a closed-form pricing formula exists. We will refer to this as the closed-form GARCH or CFG model. Assume that returns are generated by the process

$$R_{t+1} \equiv \ln(S_{t+1}) - \ln(S_t) = r_f + \lambda\sigma^2_{t+1} + \sigma_{t+1}z_{t+1}$$
$$\text{with } z_{t+1} \sim N(0,1), \text{ and } \sigma^2_{t+1} = \omega + \alpha\,(z_t - \theta\sigma_t)^2 + \beta\sigma^2_t$$

Notice that the risk premium is now multiplied by the conditional variance not standard deviation, and that z_t enters in the variance innovation term without being scaled by σ_t. Variance persistence in this model can be derived as $\alpha\theta^2 + \beta$ and the unconditional variance as $(\omega+\alpha)/(1-\alpha\theta^2-\beta)$.

The risk-neutral version of this process is

$$R_{t+1} \equiv \ln(S_{t+1}) - \ln(S_t) = r_f - \frac{1}{2}\sigma^2_{t+1} + \sigma_{t+1}z^*_{t+1}$$
$$\text{with } z^*_{t+1} \sim N(0,1), \text{ and } \sigma^2_{t+1} = \omega + \alpha\left(z^*_t - \theta^*\sigma_t\right)^2 + \beta\sigma^2_t$$

To verify that the risky assets earn the risk-free rate under the risk-neutral measure, we check again that

$$\begin{aligned}
E^*_t\left[S_{t+1}/S_t\right] &= E^*_t\left[\exp\left(r_f - \frac{1}{2}\sigma^2_{t+1} + \sigma_{t+1}z^*_{t+1}\right)\right] \\
&= \exp\left(r_f - \frac{1}{2}\sigma^2_{t+1}\right)E^*_t\left[\exp\left(\sigma_{t+1}z^*_{t+1}\right)\right] \\
&= \exp\left(r_f - \frac{1}{2}\sigma^2_{t+1}\right)\exp\left(\frac{1}{2}\sigma^2_{t+1}\right) \\
&= \exp\left(r_f\right)
\end{aligned}$$

and the variance can be verified as before as well.

Under this special GARCH process for returns, the European option price can be calculated as

$$c_{CFG} = e^{-r_f\tilde{T}}E^*_t\left[Max\left(S_{t+\tilde{T}} - X, 0\right)\right] = S_tP_1 - Xe^{-r_f\tilde{T}}P_2$$

where the formulas for P_1 and P_2 are given in the appendix. Notice that the structure of the option pricing formula is identical to that of the BSM model. As in the BSM model, P_2 is the risk-neutral probability of exercise, and P_1 is the delta of the option.

7 Implied Volatility Function (IVF) Models

The option pricing methods surveyed so far in this chapter can be derived from well-defined assumptions about the underlying dynamics of the economy. The next approach to European option pricing we consider is instead completely static and ad hoc but it turns out to offer reasonably good fit to observed option prices, and we therefore give a brief discussion of it here. The idea behind the approach is that the implied volatility smile changes only slowly over time. If we can therefore estimate a functional form on the smile today, then that functional form may work reasonably in pricing options in the near future as well.

The implied volatility smiles and smirks mentioned earlier suggest that option prices may be well captured by the following four-step approach:

1. Calculate the implied BSM volatilities for all the observed option prices on a given day as

$$\sigma_i^{iv} = c_{BSM}^{-1}\left(S_t, r_f, X_i, \tilde{T}_i, c_i^{mkt}\right) \quad \text{for } i = 1, 2, \ldots, n$$

2. Regress the implied volatilities on a second-order polynomial in moneyness and maturity. That is, use ordinary least squares (OLS) to estimate the a parameters in the regression

$$\begin{aligned}\sigma_i^{iv} = {} & a_0 + a_1 (S_t/X_i) + a_2 (S_t/X_i)^2 + a_3 \left(\tilde{T}_i/365\right) + a_4 \left(\tilde{T}_i/365\right)^2 \\ & + a_5 (S_t/X_i) \left(\tilde{T}_i/365\right) + e_i\end{aligned}$$

where e_i is an error term and where we have rescaled maturity to be in years rather than days. The rescaling is done to make the different a coefficients have roughly the same order of magnitude. This will yield the implied volatility surface as a function of moneyness and maturity. Other functional forms could of course be used.

3. Compute the fitted values of implied volatility from the regression

$$\begin{aligned}\hat{\sigma}^{iv}(S_t/X_i, \tilde{T}_i; \hat{a}) = {} & \hat{a}_0 + \hat{a}_1 (S_t/X_i) + \hat{a}_2 (S_t/X_i)^2 + \hat{a}_3 \left(\tilde{T}_i/365\right) + \hat{a}_4 \left(\tilde{T}_i/365\right)^2 \\ & + \hat{a}_5 (S_t/X_i) \left(\tilde{T}_i/365\right)\end{aligned}$$

4. Calculate model option prices using the fitted volatilities and the BSM option pricing formula, as in

$$c_{IVF} = c(S_t, r_f, X_i, \tilde{T}_i; Max(\hat{\sigma}^{iv}(S/X_i, \tilde{T}_i/365; \hat{a}), 0.0001))$$

where the $Max(\bullet)$ function ensures that the volatility used in the option pricing formula is positive.

Notice that this option pricing approach requires only a sequence of simple calculations and it is thus easily implemented.

While this four-step linear estimation approach is standard, we can typically obtain much better model option prices if the following modified estimation approach is taken. We can use a numerical optimization technique to solve for $a = \{a_0, a_1, a_2, a_3, a_4, a_5\}$ by minimizing the mean squared error

$$MSE_{MIVF} = \min_{a} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(c_i^{mkt} - c(S_t, r_f, X_i, \tilde{T}_i; Max(\sigma^{iv}(S/X_i, \tilde{T}_i/365; a), 0.01)) \right)^2 \right\}$$

The downside of this method is clearly that a numerical solution technique rather than simple OLS is needed to find the parameters. We refer to this approach as the modified implied volatility function (MIVF) technique.

8 Summary

This chapter has surveyed some key models for pricing European options. First, we introduced the simple but powerful binomial tree approach to option pricing. Then we discussed the famous Black-Scholes-Merton (BSM) model. The key assumption underlying the BSM model is that the underlying asset return dynamics are captured by the normal distribution with constant volatility. While the BSM model provides crucial insight into the pricing of derivative securities, the underlying assumptions are clearly violated by observed asset returns. We therefore next considered a generalization of the BSM model that was derived from the Gram-Charlier (GC) expansion around the normal distribution. The GC distribution allows for skewness and kurtosis and it therefore offers a more accurate description of observed returns than does the normal distribution. However, the GC model still assumes that volatility is constant over time, which we argued in earlier chapters was unrealistic. Next, we therefore presented two types of GARCH option pricing models. The first type allowed for a wide range of variance specifications, but the option price had to be calculated using Monte Carlo simulation or another numerical technique since no closed-form formula existed. The second type relied on a particular GARCH specification but in return provided a closed-form solution for the option price. Finally, we introduced the ad hoc implied volatility function (IVF) approach, which in essence consists of a second-order polynomial approximation of the implied volatility smile.

Appendix: The CFG Option Pricing Formula

The probabilities P_1 and P_2 in the closed-form GARCH (CFG) formula are derived by first solving for the conditional moment generating function. The conditional, time-t, moment generating function of the log asset prices as time $t+\tilde{T}$ is

$$f_{t,t+\tilde{T}}(\varphi) = E_t\left[\exp\left(\varphi \ln\left(S_{t+\tilde{T}}\right)\right)\right] = E_t\left[S_{t+\tilde{T}}^{\varphi}\right]$$

In the CFG model, this function takes a log-linear form (omitting the time subscripts on $f(\varphi)$)

$$f(\varphi) = S_t^{\varphi} \bullet \exp\left(A\left(t; t+\tilde{T}, \varphi\right) + B\left(t; t+\tilde{T}, \varphi\right)\sigma_{t+1}^2\right)$$

where

$$\begin{aligned} A\left(t; t+\tilde{T}, \varphi\right) = {} & A\left(t+1; t+\tilde{T}, \varphi\right) + \varphi r_f + B\left(t+1; t+\tilde{T}, \varphi\right)\omega \\ & - \frac{1}{2}\ln\left(1 - 2\alpha B\left(t+1; t+\tilde{T}, \varphi\right)\right) \end{aligned}$$

and

$$B\left(t; t+\tilde{T}, \varphi\right) = \varphi(\lambda+\theta) - \frac{1}{2}\theta^2 + \beta B\left(t+1; t+\tilde{T}, \varphi\right) + \frac{\frac{1}{2}(\varphi-\theta)^2}{1 - 2\alpha B\left(t+1; t+\tilde{T}, \varphi\right)}$$

These functions can be solved by solving backward one period at a time from the maturity date using the terminal conditions

$$A\left(t+\tilde{T}; t+\tilde{T}, \varphi\right) = 0 \text{ and } B\left(t+\tilde{T}; t+\tilde{T}, \varphi\right) = 0$$

A fundamental result in probability theory establishes the following relationship between the characteristic function $f(i\varphi)$ and the probability density function $p(x)$:

$$\int_A^{\infty} p(x)\,dx = \frac{1}{2} + \frac{1}{\pi}\int_0^{\infty} \mathrm{Re}\left[\frac{\exp(-i\varphi A) f(i\varphi)}{i\varphi}\right] d\varphi$$

where the $\mathrm{Re}(\bullet)$ function takes the real value of the argument.

Using these results, we can calculate the conditional expected payoff as

$$\begin{aligned} E_t\left[Max\left(S_{t+\tilde{T}} - X, 0\right)\right] &= E_t\left[Max\left(\exp\left(\ln\left(S_{t+\tilde{T}}\right)\right) - X, 0\right)\right] \\ &= \int_{\ln(X)}^{\infty} \exp(x)\,p(x)\,dx - X\int_{\ln(X)}^{\infty} p(x)\,dx \\ &= f(1)\left(\frac{1}{2} + \frac{1}{\pi}\int_0^{\infty}\mathrm{Re}\left[\frac{X^{-i\varphi} f(i\varphi+1)}{i\varphi f(1)}\right]d\varphi\right) \\ &\quad - X\left(\frac{1}{2} + \frac{1}{\pi}\int_0^{\infty}\mathrm{Re}\left[\frac{X^{-i\varphi} f(i\varphi)}{i\varphi}\right]d\varphi\right) \end{aligned}$$

To price the call option, we use the risk-neutral distribution to get

$$\begin{aligned}
c_{CFG} &= e^{-r_f \tilde{T}} E_t^* \left[Max\left(S_{t+\tilde{T}} - X, 0\right)\right] \\
&= S_t \left(\frac{1}{2} + \frac{1}{\pi} \int_0^\infty \mathrm{Re}\left[\frac{X^{-i\varphi} f^*(i\varphi + 1)}{i\varphi f^*(1)} \right] d\varphi \right) \\
&\quad - Xe^{-r_f \tilde{T}} \left(\frac{1}{2} + \frac{1}{\pi} \int_0^\infty \mathrm{Re}\left[\frac{X^{-i\varphi} f^*(i\varphi)}{i\varphi} \right] d\varphi \right) \\
&\equiv S_t P_1 - Xe^{-r\tilde{T}} P_2
\end{aligned}$$

where we have used the fact that $f^*(1) = E_t^*\left[S_{t+\tilde{T}}\right] = e^{r_f \tilde{T}} S_t$. Note that under the risk-neutral distribution, λ is set to $-\frac{1}{2}$, and θ is replaced by θ^*. Finally, we note that the previous integrals must be solved numerically.

Further Resources

This chapter has focused on option pricing in discrete time in order to remain consistent with the previous chapters. There are many excellent textbooks on options. The binomial tree model in this chapter follows Hull (2011) who also provides a proof that the binomial model converges to the BSM model when the number of steps in the tree goes to infinity.

The classic papers on the BSM model are Black and Scholes (1973) and Merton (1973). The discrete time derivations in this chapter were introduced in Rubenstein (1976) and Brennan (1979). Merton (1976) introduced a continuous time diffusion model with jumps allowing for kurtosis in the distribution of returns. See Andersen and Andreasen (2000) for extensions to Merton's 1976 model.

For recent surveys on empirical option valuation, see Bates (2003), Garcia et al. (2010), and Christoffersen et al. (2010a).

The GC model is derived in Backus et al. (1997). The general GARCH option pricing framework is introduced in Duan (1995). Duan and Simonato (1998) discuss Monte Carlo simulation techniques for the GARCH model and Duan et al. (1999) contains an analytical approximation to the GARCH model price. Ritchken and Trevor (1999) suggest a trinomial tree method for calculating the GARCH option price.

Duan (1999) and Christoffersen et al. (2010b) consider extensions to the GARCH option pricing model allowing for conditionally nonnormal returns. The closed-form GARCH option pricing model is derived in Heston and Nandi (2000) and extended in Christoffersen et al. (2006).

Christoffersen and Jacobs (2004a) compared the empirical performance of various GARCH variance specifications for option pricing and found that the simple variance specification including a leverage effect as applied in this chapter works very well compared with the BSM model.

Hsieh and Ritchken (2005) compared the GARCH (GH) and the closed-form GARCH (CFG) models and found that the GH model performs the best in terms of out-of-sample option valuation. See Christoffersen et al. (2008) and Christoffersen et al. (2010a) for option valuation using the GARCH component models described in Chapter 4.

GARCH option valuation models with jumps in the innovations have been developed by Christoffersen et al. (2011) and Ornthanalai (2011).

Hull and White (1987) and Heston (1993) derived continuous time option pricing models with time-varying volatility. Bakshi et al. (1997) contains an empirical comparison of Heston's model with more general models and finds that allowing for time-varying volatility is key in fitting observed option prices. Lewis (2000) discusses the implementation of option valuation models with time-varying volatility.

The IVF model is described in Dumas et al. (1998) and the modified IVF model (MIVF) is examined in Christoffersen and Jacobs (2004b), who find that the MIVF model performs very well empirically compared with the simple BSM model. Berkowitz (2010) provides a theoretical justification for the MIVF approach. Bams et al. (2009) discuss the choice of objective function in option model calibration.

References

Andersen, L., Andreasen, J., 2000. Jump-diffusion processes: Volatility smile fitting and numerical methods for option pricing. Rev. Derivatives Res. 4, 231–262.

Backus, D., Foresi, S., Li, K., Wu, L., 1997. Accounting for Biases in Black-Scholes. Manuscript, The Stern School at New York University.

Bakshi, G., Cao, C., Chen, Z., 1997. Empirical performance of alternative option pricing models. J. Finance 52, 2003–2050.

Bams, D., Lehnert, T., Wolff, C., 2009. Loss functions in option valuation: A framework for selection. Manag. Sci. 55, 853–862.

Bates, D., 2003. Empirical option pricing: A retrospection. J. Econom. 116, 387–404.

Berkowitz, J., 2010. On justifications for the ad hoc Black-Scholes method of option pricing. Stud. Nonlinear Dyn. Econom. 14 (1), Article 4.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. J. Polit. Econ. 81, 637–659.

Brennan, M., 1979. The pricing of contingent claims in discrete time models. J. Finance 34, 53–68.

Christoffersen, P., Dorion, C., Jacobs, K., Wang, Y., 2010a. Volatility components: Affine restrictions and non-normal innovations. J. Bus. Econ. Stat. 28, 483–502.

Christoffersen, P., Elkamhi, R., Feunou, B., Jacobs, K., 2010b. Option valuation with conditional heteroskedasticity and non-normality. Rev. Financ. Stud. 23, 2139–2183.

Christoffersen, P., Heston, S., Jacobs, K., 2006. Option valuation with conditional skewness. J. Econom. 131, 253–284.

Christoffersen, P., Jacobs, K., 2004a. Which GARCH model for option valuation? Manag. Sci. 50, 1204–1221.

Christoffersen, P., Jacobs, K., 2004b. The importance of the loss function in option valuation. J. Financ. Econ. 72, 291–318.

Christoffersen, P., Jacobs, K., Ornthanalai, C., 2010. GARCH option valuation, theory and evidence. In: Duan, J.-C., Gentle, J., Hardle, W. (Eds.), Handbook of Computational Finance, forthcoming. Springer, New York, NY.

Christoffersen, P., Jacobs, K., Ornthanalai, C., 2011. Exploring time-varying jump intensities: Evidence from S&P 500 returns and options. Available from: SSRN, http://ssrn.com/abstract=1101733.

Christoffersen, P., Jacobs, K., Ornthanalai, C., Wang, Y., 2008. Option valuation with long-run and short-run volatility components. J. Financ. Econ. 90, 272–297.

Duan, J., 1995. The GARCH option pricing model. Math. Finance 5, 13–32.

Duan, J., 1999. Conditionally Fat-Tailed Distributions and the Volatility Smile in Options. Manuscript, Hong Kong University of Science and Technology.

Duan, J., Gauthier, G., Simonato, J.-G., 1999. An analytical approximation for the GARCH option pricing model. J. Comput. Finance 2, 75–116.

Duan, J., Simonato, J.-G., 1998. Empirical martingale simulation for asset prices. Manag. Sci. 44, 1218–1233.

Dumas, B., Fleming, J., Whaley, R., 1998. Implied volatility functions: Empirical tests. J. Finance 53, 2059–2106.

Garcia, R., Ghysels, E., Renault, E., 2010. The econometrics of option pricing. In: Aït-Sahalia, Y., Hansen, L.P. (Eds.), Handbook of Financial Econometrics. Elsevier, North Holland, pp. 479–552.

Heston, S., 1993. A closed-form solution for options with stochastic volatility, with applications to bond and currency options. Rev. Financ. Stud. 6, 327–343.

Heston, S., Nandi, S., 2000. A closed-form GARCH option pricing model. Rev. Financ. Stud. 13, 585–626.

Hsieh, K., Ritchken, P., 2005. An empirical comparison of GARCH option pricing models. Rev. Derivatives Res. 8, 129–150.

Hull, J., 2011. Options, Futures and Other Derivatives, eighth ed. Prentice-Hall, Upper Saddle River, NJ.

Hull, J., White, A., 1987. The pricing of options on assets with stochastic volatilities. J. Finance 42, 281–300.

Lewis, A., 2000. Option Valuation under Stochastic Volatility. Finance Press, Newport Beach, California.

Merton, R., 1973. Theory of rational option pricing. Bell J. Econ. Manag. Sci. 4, 141–183.

Merton, R., 1976. Option pricing when underlying stock returns are discontinuous. J. Financ. Econ. 3, 125–144.

Ornthanalai, C., 2011. A new class of asset pricing models with Levy processes: Theory and applications. Available from: SSRN, http://ssrn.com/abstract=1267432.

Ritchken, P., Trevor, R., 1999. Pricing options under generalized GARCH and stochastic volatility processes. J. Finance 54, 377–402.

Rubenstein, M., 1976. The valuation of uncertain income streams and the pricing of options. Bell J. Econ. Manag. Sci. 7, 407–425.

Empirical Exercises

Open the Chapter10Data.xlsx file from the web site. The file contains European call options on the S&P 500 from January 6, 2010.

1. Calculate the BSM price for each option using a standard deviation of 0.01 per day. Using Solver, find the volatility that minimizes the mean squared pricing error using 0.01 as a starting value. Keep the BSM prices that correspond to this optimal volatility and use these prices below.
2. Scatter plot the BSM pricing errors (actual price less model price) against moneyness defined as (S/X) for the different maturities.

3. Calculate the implied BSM volatility (standard deviation) for each of the options. You can use Excel's Solver to do this. Scatter plot the implied volatilities against moneyness.
4. Fit the Gram-Charlier option price to the data. Estimate a model with skewness only. Use nonlinear test squares (NLS).
5. Regress implied volatility on a constant, moneyness, the time-to-maturity divided by 365, each variable squared, and their cross product. Calculate the fitted BSM volatility from the regression for each option. Calculate the ad hoc IVF price for each option using the fitted values for volatility.
6. Redo the IVF estimation using NLS to minimize the mean squared pricing error (MSE). Call this MIVF. Use the IVF regression coefficients as starting values.
7. Calculate the square root of the mean squared pricing error from the IVF and MIVF models and compare them to the square root of the MSE from the standard BSM model and the Gram-Charlier model. Scatter plot the pricing errors from the MIVF model against moneyness and compare them to the plots from exercise 2.

The answers to these exercises can be found in the Chapter10Results.xlsx file on the companion site.

For more information see the companion site at
http://www.elsevierdirect.com/companions/9780123744487

11 Option Risk Management

1 Chapter Overview

In the previous chapter, we gave a brief overview of various models for pricing options. In this chapter, we turn our attention to the key task of incorporating derivative securities into the portfolio risk model, which we developed in previous chapters. Just as the nonlinear payoff function was the key feature from the perspective of option pricing in the previous chapter, it is also driving the risk management discussion in this chapter. The nonlinear payoff creates asymmetry in the portfolio return distribution, even if the return on the underlying asset follows a symmetric distribution. Getting a handle on this asymmetry is a key theme of this chapter.

The chapter is structured as follows:

- We define the delta of an option, which provides a linear approximation to the nonlinear option price. We then present delta formulas from the various models introduced in the previous chapter.
- We establish the delta-based approach to portfolio risk management. The idea behind this approach is to linearize the option return and thereby make it fit into the risk models discussed earlier in the book. The downside of this approach is that it ignores the key asymmetry in option payoffs.
- We define the gamma of an option, which gives a second-order approximation of the option price as a function of the underlying asset price.
- We use the gamma of an option to construct a quadratic model of the portfolio return distribution. We discuss two implementations of the quadratic model: one relies on the Cornish-Fisher approximation from Chapter 6, and the other relies on the Monte Carlo simulation technique from Chapter 8.
- We will measure the risk of options using the full valuation method, which relies on an accurate but computationally intensive version of the Monte Carlo simulation technique from Chapter 8.
- We illustrate all the suggested methods in a simple example. We then discuss a major pitfall in the use of the linear and quadratic approximations in another numerical example. This pitfall, in turn, motivates the use of the full valuation model.

Elements of Financial Risk Management. DOI: 10.1016/B978-0-12-374448-7.00011-7

2 The Option Delta

The delta of an option is defined as the partial derivative of the option price with respect to the underlying asset price, S_t. For puts and calls, we define

$$\delta^c \equiv \frac{\partial c}{\partial S_t}$$
$$\delta^p \equiv \frac{\partial p}{\partial S_t}$$

Notice that the deltas are not observed in the market, but instead are based on the assumed option pricing model.

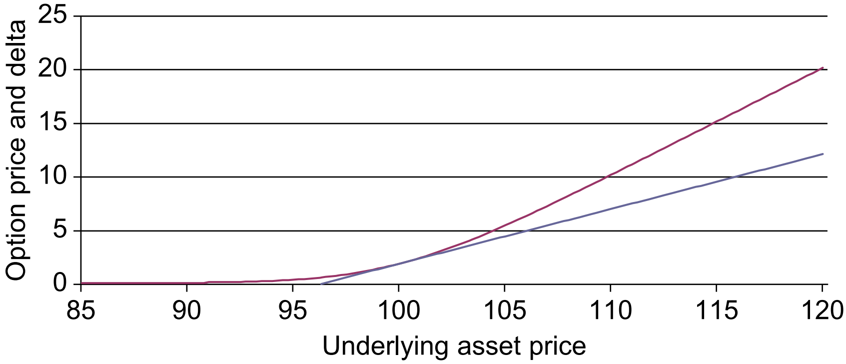
Figure 11.1 illustrates the familiar tangent interpretation of a partial derivative. The option price for a generic underlying asset price, S, is approximated by

$$c(S) \approx c(S_t) + \delta(S - S_t)$$

where S_t is the current price of the underlying asset. In Figure 11.1, S_t equals 100.

The delta of an option (in this case, a call option) can be viewed as providing a linear approximation (around the current asset price) to the nonlinear option price, where the approximation is reasonably good for asset prices close to the current price but gets gradually worse for prices that deviate significantly from the current price, as Figure 11.1 illustrates. To a risk manager, the poor approximation of delta to the true option price for large underlying price changes is clearly unsettling. Risk management is all about large price changes, and we will therefore consider more accurate approximations here.

Figure 11.1 Call option price and delta approximation.



Notes: The call option price (red) and the delta approximation (blue) of the option are plotted against the price of the underlying asset. The strike price is 100 and delta is calculated at an asset price of 100.

2.1 The Black-Scholes-Merton Model

Recall, from the previous chapter, the Black-Scholes-Merton (BSM) formula for a European call option price

$$c_{BSM} = S_t\Phi(d) - \exp(-r_f\tilde{T})X\Phi(d - \sigma\sqrt{\tilde{T}})$$

where $\Phi(\bullet)$ is the cumulative density of a standard normal variable, and

$$d = \frac{\ln(S_t/X) + \tilde{T}\left(r_f + \sigma^2/2\right)}{\sigma\sqrt{\tilde{T}}}$$

Using basic calculus, we can take the partial derivative of the option price with respect to the underlying asset price, S_t, as follows:

$$\frac{\partial c_{BSM}}{\partial S_t} \equiv \delta^c_{BSM} = \Phi(d)$$

We refer to this as the delta of the option, and it has the interpretation that for small changes in S_t the call option price will change by $\Phi(d)$. Notice that as $\Phi(\bullet)$ is the normal cumulative density function, which is between zero and one, we have

$$0 < \delta^c_{BSM} < 1$$

so that the call option price in the BSM model will change in the same direction as the underlying asset price, but the change will be less than one-for-one.

For a European put option, we have the put-call parity stating that

$$\begin{aligned} S_t + p &= c + X\exp(-r_f\tilde{T}), \text{ or} \\ p &= c + X\exp(-r_f\tilde{T}) - S_t \end{aligned}$$

so that we can easily derive

$$\frac{\partial p_{BSM}}{\partial S_t} \equiv \delta^p_{BSM} = \frac{\partial c_{BSM}}{\partial S_t} - 1 = \Phi(d) - 1$$

Notice that we have

$$-1 < \delta^p_{BSM} < 0$$

so that the BSM put option price moves in the opposite direction of the underlying asset, and again the option price will change by less (in absolute terms) than the underlying asset price.

In the case where a dividend or interest is paid on the underlying asset at a rate of q per day, the deltas will be

$$\begin{aligned} \delta^c_{BSM} &= \exp(-q\tilde{T})\Phi(d), \\ \delta^p_{BSM} &= \exp(-q\tilde{T})\left(\Phi(d) - 1\right) \end{aligned}$$

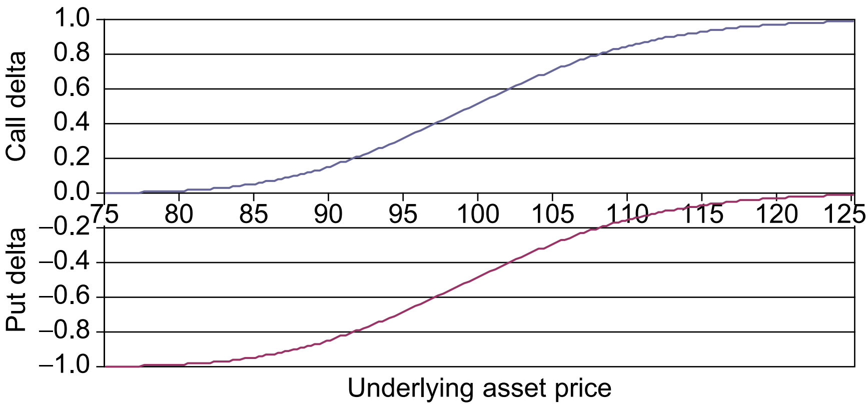
where

$$d = \frac{\ln(S_t/X) + \tilde{T}\left(r_f - q + \sigma^2/2\right)}{\sigma\sqrt{\tilde{T}}}$$

The deltas of the European call and put options from the BSM model are shown in Figure 11.2 for $X = 100$ and for S_t varying from 75 to 125. Notice that delta changes most dramatically when the option is close to at-the-money—that is, when $S_t \approx X$. A risk management model that relies on a fixed initial delta is therefore likely to be misleading when the portfolio contains a significant number of at-the-money options.

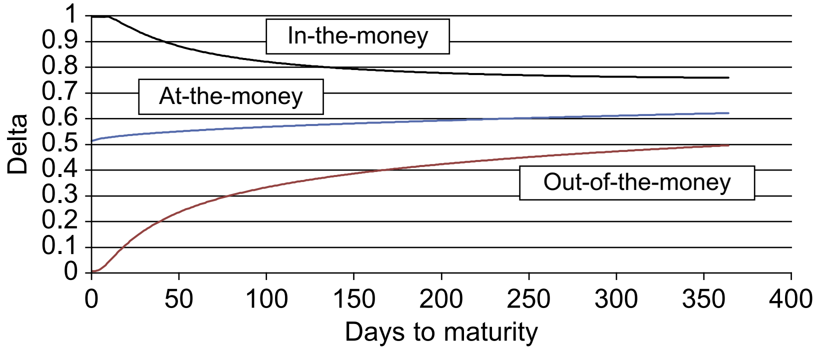
Figure 11.3 shows the delta of three call options with different strike prices ($X = 80$, 100, and 120, respectively) plotted against maturity, $\tilde{T}$, ranging from 1 to 365

Figure 11.2 The delta of a call option (top) and a put option (bottom).



Notes: The plot shows the delta of a call option (blue) and a put option (red) as a function of the underlying asset price. The strike price is 100 for both options.

Figure 11.3 The delta of three call options.



Notes: We plot the delta of three call options with different strike prices. Moving from right to left in the plot the maturity of each option goes from one year to one day.

calendar days on the horizontal axis. The asset price S_t is held fixed at 100 throughout the graph.

Notice when the maturity gets shorter (we move from right to left in the graph), the deltas diverge: the delta from the in-the-money call option increases to 1, the delta from the out-of-the-money option decreases to 0, and the delta from the at-the-money option converges to 0.5. An in-the-money call option with short maturity is almost certain to pay off $S_t - X$, which is why its price moves in tandem with the asset price S_t and its delta is close to 1. An out-of-the-money option with short maturity is almost certain to pay 0, so that its price is virtually constant and its delta is close to 0.

2.2 The Binomial Tree Model

Option deltas can also be computed using binomial trees. This is particularly important for American put options for which early exercise may be optimal, which will impact the current option price and also the option delta. Table 11.1 shows the binomial tree example from Table 10.4 in Chapter 10. The black font again shows the American put option price at each node. The green font now shows the option delta.

The delta at point A (that is at present) can be computed very easily in binomial trees simply as

$$\delta_{Bin} = \frac{Put_B - Put_C}{S_B - S_C}$$

Table 11.1 Delta of American put option

Market Variables				
$S_t =$	1000			D
Annual $r_f =$	0.05			1528.47
Contract Terms				0.00
$X =$	1100			
$T =$	0.25		B	
			1236.31	
Parameters			−0.19	
Annual Vol =	0.6		53.48	
tree steps =	2			
$dt =$	0.125	A		E
$u =$	1.23631111	1000.00		1000.00
$d =$	0.808857893	−0.56		
RNP =	0.461832245	180.25		100.00
Stock is black			C	
American put delta is green			808.86	
American put price is red			−1.00	
			291.14	
				F
				654.25
				445.75

Notes: The green font shows the delta of the American put option at points A, B, and C in the tree.

and a similar formula can be used for European puts as well as for call options of each style. Note that delta was already used in Chapter 10 to identify the number of units in the underlying asset we needed to buy to hedge the sale of one option. Delta changes in each point of the tree, which shows that option positions require dynamic hedging in order to remain risk free.

2.3 The Gram-Charlier Model

As the delta is a partial derivative of an option pricing model with respect to the underlying asset price, it is fundamentally model dependent. The preceding deltas were derived from the BSM model, but different option pricing models imply different formulas for the deltas. We saw in the previous chapter that the BSM model sometimes misprices traded options quite severely. We therefore want to consider using more accurate option pricing models for calculating the options delta.

In the case of the Gram-Charlier option pricing model, we have

$$\begin{aligned} c_{GC} &\approx S_t\Phi(d) - Xe^{-r\tilde{T}}\Phi\left(d-\sqrt{\tilde{T}}\sigma\right) \\ &\quad + S_t\phi(d)\sigma\left[\frac{\zeta_{11}}{3!}\left(2\sqrt{\tilde{T}}\sigma - d\right) - \frac{\zeta_{21}/\sqrt{\tilde{T}}}{4!}\left(1-d^2+3d\sqrt{\tilde{T}}\sigma - 3\tilde{T}\sigma^2\right)\right] \end{aligned}$$

and the partial derivative with respect to the asset price in this case is

$$\begin{aligned} \delta_{GC} &= \frac{\partial c_{GC}}{\partial S_t} = \Phi(d) - \frac{\zeta_{11}/\sqrt{\tilde{T}}}{3!}\phi(d)\left(1-d^2+3\sigma\sqrt{\tilde{T}}d - 2\sigma^2\tilde{T}\right) \\ &\quad + \frac{\zeta_{21}/\tilde{T}}{4!}\phi(d)\left[3d\left(1-2\sigma^2\tilde{T}\right) + 4d^2\sigma\sqrt{\tilde{T}} - d^3 - 4\sigma\sqrt{\tilde{T}} + 3\sigma^3\tilde{T}^{3/2}\right] \end{aligned}$$

which collapses to the BSM delta of $\Phi(d)$ when skewness, ζ_{11}, and excess kurtosis, ζ_{21}, are both zero. Again, we can easily calculate the put option delta from

$$\delta^p_{GC} \equiv \frac{\partial p_{GC}}{\partial S_t} = \frac{\partial c_{GC}}{\partial S_t} - 1$$

2.4 The GARCH Option Pricing Models

Calculating deltas from the general GARCH option pricing model, we are immediately faced with the issue that the option price is not available in closed form but must be simulated. We have in general

$$c_{GH} = \exp(-r_f\tilde{T})E_t^*\left[Max\left\{S_{t+\tilde{T}} - X, 0\right\}\right]$$

which we compute by simulation as

$$c_{GH} \approx \exp(-r_f\tilde{T})\frac{1}{MC}\sum_{i=1}^{MC} Max\left\{\check{S}^*_{i,t+\tilde{T}} - X, 0\right\}$$

where $\check{S}^*_{i,t+\tilde{T}}$ is the hypothetical GARCH asset price on option maturity date $t+\tilde{T}$ for Monte Carlo simulation path i, where the simulation is done under the risk-neutral distribution.

The partial derivative of the GARCH option price with respect to the underlying asset price can be shown to be

$$\delta^c_{GH} = \exp(-r_f\tilde{T})E^*_t\left[\frac{S_{t+\tilde{T}}}{S_t}\mathbf{1}\left(S_{t+\tilde{T}} \geq X\right)\right]$$

where the function $\mathbf{1}(\bullet)$ takes the value 1 if the argument is true and zero otherwise.

The GARCH delta must also be found by simulation as

$$\delta^c_{GH} \approx \exp(-r_f\tilde{T})\frac{1}{MC}\sum_{i=1}^{MC}\frac{\check{S}^*_{i,t+\tilde{T}}}{S_t}\mathbf{1}\left(\check{S}^*_{i,t+\tilde{T}} \geq X\right)$$

where $\check{S}^*_{i,t+\tilde{T}}$ is again the simulated future risk-neutral asset price. The delta of the European put option can still be derived from the put-call parity formula.

In the special case of the closed-form GARCH process, we have the European call option pricing formula

$$c_{CFG} = S_tP_1 - Xe^{-r_f\tilde{T}}P_2$$

and the delta of the call option is

$$\delta^c_{CFG} = P_1$$

The formula for P_1 is given in the appendix to the previous chapter.

3 Portfolio Risk Using Delta

Equipped with a formula for delta from our option pricing formula of choice, we are now ready to adapt our portfolio distribution model from earlier chapters to include portfolios of options.

Consider a portfolio consisting of just one (long) call option on a stock. The change in the dollar value (or the dollar return) of the option portfolio, $DV_{PF,t+1}$, is then just the change in the value of the option

$$DV_{PF,t+1} \equiv c_{t+1} - c_t$$

Using the delta of the option, we have that for small changes in the underlying asset price

$$\delta \approx \frac{c_{t+1} - c_t}{S_{t+1} - S_t}$$

Defining geometric returns on the underlying stock as

$$r_{t+1} = \frac{S_{t+1} - S_t}{S_t} \approx \ln\left(S_{t+1}/S_t\right) = R_{t+1}$$

and combining the previous three equations, we get the change in the option portfolio value to be

$$DV_{PF,t+1} \approx \delta\left(S_{t+1} - S_t\right) \approx \delta S_t R_{t+1}$$

The upshot of this formula is that we can write the change in the dollar value of the option as a known value δ_t times the future return of the underlying asset, R_{t+1}, if we rely on the delta approximation to the option pricing formula.

Notice that a portfolio consisting of an option on a stock corresponds to a stock portfolio with δ shares. Similarly, we can think of holdings in the underlying asset as having a delta of 1 per share of the underlying asset. Trivially, the derivative of a stock price with respect to the stock price is 1. Thus, holding one share corresponds to having $\delta = 1$, and holding 100 shares corresponds to having $\delta = 100$.

Similarly, a short position of 10 identical calls corresponds to setting $\delta = -10\delta^c$, where δ^c is the delta of each call option. The delta of a short position in call options is negative, and the delta of a short position in put options is positive as the delta of a put option itself is negative.

The variance of the portfolio in the delta-based model is

$$\sigma^2_{DV,t+1} \approx \delta^2 S_t^2 \sigma^2_{t+1}$$

where σ^2_{t+1} is the conditional variance of the return on the underlying stock.

Assuming conditional normality, the dollar Value-at-Risk (*VaR*) in this case is

$$\$VaR^p_{t+1} = -\sigma_{DV,t+1}\Phi_p^{-1} \approx -abs\left(\delta\right)S_t\sigma_{t+1}\Phi_p^{-1}$$

where the absolute value, $abs(*)$, comes from having taken the square root of the portfolio change variance, $\sigma^2_{DV,t+1}$. Notice that since $DV_{PF,t+1}$ is measured in dollars, we are calculating dollar *VaR*s directly and not percentage *VaR*s as in previous chapters. The percentage *VaR* can be calculated immediately from the dollar *VaR* by dividing it by the current value of the portfolio.

In case we are holding a portfolio of several options on the *same* underlying asset, we can simply add up the deltas. The delta of a portfolio of options on the same underlying asset is just the weighted sum of the individual deltas as in

$$\delta = \sum_j m_j \delta_j$$

where the weight, m_j, equals the number of the particular option contract j. A short position in a particular type of options corresponds to a negative m_j.

In the general case where the portfolio consists of options on n underlying assets, we have

$$DV_{PF,t+1} \approx \sum_{i=1}^{n} \delta_i S_{i,t} R_{i,t+1}$$

In this delta-based model, the variance of the dollar change in the portfolio value is again

$$\sigma_{DV,t+1}^2 \approx \sum_{i=1}^{n}\sum_{j=1}^{n} abs\left(\delta_i\right) abs(\delta_j)\, S_{i,t} S_{j,t} \sigma_{ij,t+1}$$

Under conditional normality, the dollar VaR of the portfolio is again just

$$\$VaR_{t+1}^{p} = -\sigma_{DV,t+1}\Phi_p^{-1}$$

Thus, in this case, we can use the Gaussian risk management framework established in Chapters 4 and 7 without modification. The linearization of the option prices through the use of delta, together with the assumption of normality, makes the calculation of the VaR and other risk measures very easy.

Notice that if we allow for the standard deviations, $\sigma_{i,t+1}$, to be time varying as in GARCH, then the option deltas should ideally be calculated from the GARCH model also. We recall that for horizons beyond one day, the GARCH returns are no longer normal, in which case the return distribution must be simulated. We will discuss simulation-based approaches to option risk management later. When volatility is assumed to be constant and returns are assumed to be normally distributed, we can calculate the dollar VaR at horizon K by

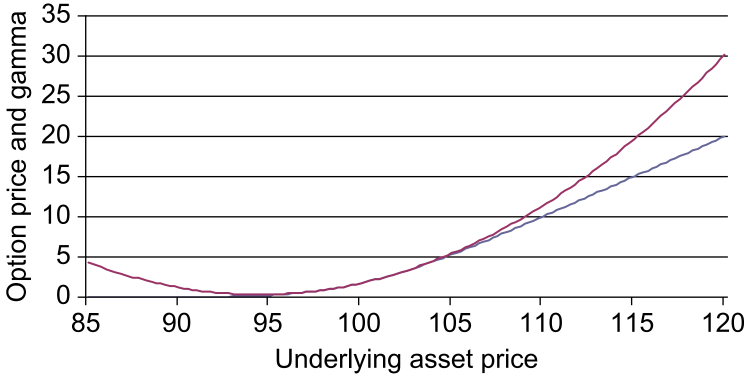
$$\$VaR_{t+1:t+K}^{p} = -\sigma_{DV}\sqrt{K}\Phi_p^{-1}$$

where σ_{DV} is the daily portfolio volatility and where K is the risk management horizon measured in trading days.

4 The Option Gamma

The linearization of the option price using the delta approach outlined here often does not offer a sufficiently accurate description of the risk from the option. When the underlying asset price makes a large upward move in a short time, the call option price will increase by more than the delta approximation would suggest. Figure 11.1 illustrates this point. If the underlying price today is \$100 and it moves to \$115, then the nonlinear option price increase is substantially larger than the linear increase in the delta approximation. Risk managers, of course, care deeply about large moves in asset prices and this shortcoming of the delta approximation is therefore a serious issue. A possible solution to this problem is to apply a quadratic rather than just a linear approximation to the option price. The quadratic approximation attempts to accommodate part of the error made by the linear delta approximation.

Figure 11.4 Call option price (blue) and the gamma approximation (red).



Notes: The call option price and the gamma approximation are plotted against the price of the underlying asset. The strike price is 100 and delta is calculated with an asset price of 100 as well.

The Greek letter gamma, γ, is used to denote the rate of change of δ with respect to the price of the underlying asset, that is,

$$\gamma \equiv \frac{\partial \delta}{\partial S_t} = \frac{\partial^2 c}{\partial S_t^2}$$

Figure 11.4 shows a call option price as a function of the underlying asset price.

The gamma approximation is shown along with the model option price. The model option price is approximated by the second-order Taylor expansion

$$c(S) \approx c(S_t) + \delta (S - S_t) + \frac{1}{2}\gamma (S - S_t)^2$$

For a European call or put on an underlying asset paying a cash flow at the rate q, and relying on the BSM model, the gamma can be derived as

$$\gamma^c = \gamma^p = \frac{\phi(d) \exp\left(-q\tilde{T}\right)}{S_t \sigma \sqrt{\tilde{T}}}, \text{ where}$$

$$d = \frac{\ln(S_t/X) + \tilde{T}\left(r_f - q + \sigma^2/2\right)}{\sigma\sqrt{\tilde{T}}}$$

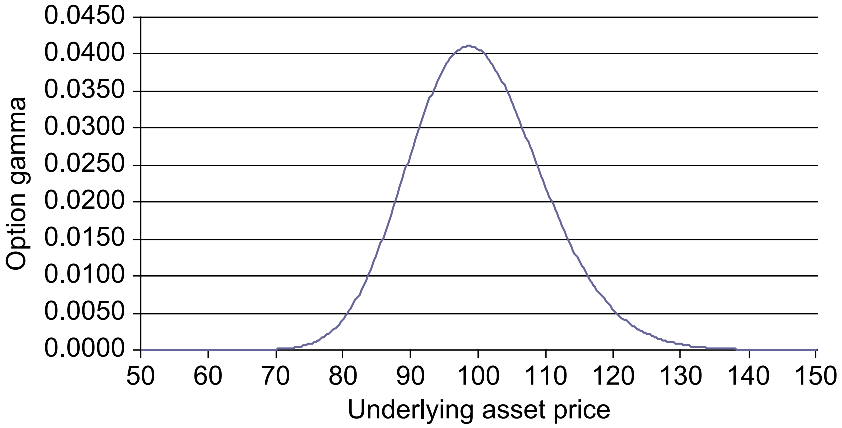
and where $\phi(\bullet)$ as before is the probability density function for a standard normal variable,

$$\phi(d) \equiv \frac{1}{\sqrt{2\pi}} \exp\left(-d^2/2\right)$$

Figure 11.5 shows the gamma for an option using the BSM model with parameters as in Figure 11.2 where we plotted the deltas.

When the option is close to at-the-money, the gamma is relatively large and when the option is deep out-of-the-money or deep in-the-money the gamma is relatively

Figure 11.5 The gamma of an option.



Notes: The figure shows the gamma of an option with a strike price of 100 plotted against the price of the underlying asset.

small. This is because the nonlinearity of the option price is highest when the option is close to at-the-money. Deep in-the-money call option prices move virtually one-for-one with the price of the underlying asset because the options will almost surely be exercised. Deep out-of-the-money options will almost surely not be exercised, and they are therefore virtually worthless regardless of changes in the underlying asset price.

All this, in turn, implies that for European options, ignoring gamma is most crucial for at-the-money options. For these options, the linear delta-based model can be highly misleading.

Finally, we note that gamma can be computed using binomial trees as well. Table 11.2 shows the gamma for the American put option from Table 11.1.

The formula used for gamma in the tree is simply

$$\gamma_{Bin} = \frac{\delta_B - \delta_C}{0.5\left[(S_D - S_E) + (S_E - S_F)\right]} = \frac{\delta_B - \delta_C}{0.5\,(S_D - S_F)}$$

and it is thus based on the change in the delta from point B to C in the tree divided by the average change in the stock price when going from points B and C.

5 Portfolio Risk Using Gamma

In the previous delta-based model, when considering a portfolio consisting of options on one underlying asset, we have

$$DV_{PF,t+1} \approx \delta S_t R_{t+1}$$

where δ denotes the weighted sum of the deltas on all the individual options in the portfolio.

Table 11.2 Gamma of American put option

Market Variables				
$S_t =$	1000			D
Annual $r_f =$	0.05			1528.47
Contract Terms				0.00
$X =$	1100			
$T =$	0.25		B	
			1236.31	
Parameters			−0.19	
Annual Vol =	0.6		53.48	
tree steps =	2			
$dt =$	0.125	A		E
$u =$	1.23631111	1000.00		1000.00
$d =$	0.808857893	−0.56		
RNP =	0.461832245	180.25		100.00
Stock is black		0.001855	C	
American put delta is green			808.86	
American put price is red			−1.00	
American put gamma is blue			291.14	
				F
				654.25
				445.75

Notes: The blue font shows the gamma of an American put option at point A computed using the deltas at points B and C.

When incorporating the second derivative, gamma, we instead rely on the quadratic approximation

$$DV_{PF,t+1} \approx \delta S_t R_{t+1} + \frac{1}{2}\gamma S_t^2 R_{t+1}^2$$

where the portfolio δ and γ are calculated as

$$\delta = \sum_j m_j \delta_j$$

$$\gamma = \sum_j m_j \gamma_j$$

where again m_j denotes the number of option contract j in the portfolio.

5.1 The Cornish-Fisher Approximation

If we assume that the underlying asset return, R_{t+1}, is normally distributed with mean zero and constant variance σ^2, and rely on the preceding quadratic approximation, then the first three moments of the distribution of changes in the value of a portfolio of options can be written as

$$\mu_{DV} \approx \frac{1}{2}\gamma S_t^2 \sigma^2$$

$$\sigma_{DV}^2 \approx \delta^2 S_t^2 \sigma^2 + \frac{1}{2}\gamma^2 S_t^4 \sigma^4$$

$$\zeta_{1,DV} \approx \frac{\frac{9}{2}\delta^2\gamma S_t^4\sigma^4 + \frac{15}{8}\gamma^3 S_t^6\sigma^6 - 3\left(\delta^2 S_t^2\sigma^2 + \frac{3}{4}\gamma^2 S_t^4\sigma^4\right)\mu_{DV} + 2\mu_{DV}^3}{\sigma_{DV}^3}$$

For example, we can derive the expected value as

$$\mu_{DV} \equiv E\left[DV_{PF,t+1}\right] \approx E\left[\delta S_t R_{t+1}\right] + E\left[\frac{1}{2}\gamma S_t^2 R_{t+1}^2\right]$$
$$= \delta S_t \cdot 0 + \frac{1}{2}\gamma S_t^2\sigma^2 = \frac{1}{2}\gamma S_t^2 \sigma^2$$

The K-day horizon moments can be calculated by scaling σ by $\sqrt{K}$ everywhere.

Notice that because the change in the portfolio value now depends on the squares of the individual returns, the portfolio return is no longer normally distributed, even if the underlying asset return is normally distributed. In particular, we notice that even if the underlying return has mean zero, the portfolio mean is no longer zero. More important, the variance formula changes and the portfolio skewness is no longer zero, even if the underlying asset has no skewness. The asymmetry of the options payoff itself creates asymmetry in the portfolio distribution. The linear-normal model presented earlier fails to capture this skewness, but the quadratic model considered here captures the skewness at least approximately. In this way, the quadratic model can offer a distinct improvement over the linear model.

The approximate Value-at-Risk of the portfolio can be calculated using the Cornish-Fisher approach discussed in Chapter 6. The Cornish-Fisher *VaR* allowing for skewness is

$$\$VaR_{t+1}^{P} = -\mu_{DV} - \left(\Phi_p^{-1} + \frac{1}{6}\left(\left(\Phi_p^{-1}\right)^2 - 1\right)\zeta_{1,DV}\right)\sigma_{DV}$$

Unfortunately, the analytical formulas for the moments of options portfolios with many underlying assets are quite cumbersome, and they rely on the unrealistic assumption of normality and constant variance. We will therefore now consider a much more general but simulation-based technique that builds on the Monte Carlo method introduced in Chapter 8. Later, we will illustrate the Cornish-Fisher quadratic *VaR* in a numerical example.

5.2 The Simulation-Based Gamma Approximation

Consider again the simple case where the portfolio consists of options on only one underlying asset and we are interested in the K-day $\$VaR$. We have

$$DV_{PF,t+K} \approx \delta S_t R_{t+1:t+K} + \frac{1}{2}\gamma S_t^2 R_{t+1:t+K}^2$$

Using the assumed model for the physical distribution of the underlying asset return, we can simulate MC pseudo K-day returns on the underlying asset

$$\left\{\widehat{R}_h^K\right\}_{h=1}^{MC}$$

and calculate the hypothetical changes in the portfolio value as

$$\widehat{DV}_{PF,h}^K \approx \delta S_t \widehat{R}_h^K + \frac{1}{2}\gamma S_t^2 \left(\widehat{R}_h^K\right)^2$$

from which we can calculate the Value-at-Risk as

$$\$VaR_{t+1:t+K}^p = -Percentile\left\{\left\{\widehat{DV}_{PF,h}^K\right\}_{h=1}^{MC}, 100p\right\}$$

In the general case of options on n underlying assets, we have

$$DV_{PF,t+K} \approx \sum_{i=1}^{n} \delta_i S_{i,t} R_{i,t+1:t+K} + \sum_{i=1}^{n} \frac{1}{2}\gamma_i S_{i,t}^2 R_{i,t+1:t+K}^2$$

where δ_i and γ_i are the aggregate delta and gamma of the portfolio with respect to the ith return.

If we in addition allow for derivatives that depend on several underlying assets, then we write

$$DV_{PF,t+K} \approx \sum_{i=1}^{n} \delta_i S_{i,t} R_{i,t+1:t+K} + \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{1}{2}\gamma_{ij} S_{i,t} S_{j,t} R_{i,t+1:t+K} R_{j,t+1:t+K}$$

which includes the so-called cross-gammas, γ_{ij}. For a call option, for example, we have

$$\gamma_{ij}^c \equiv \frac{\partial^2 c}{\partial S_i \partial S_j}, \quad \text{for } i \neq j$$

Cross-gammas are relevant for options with multiple sources of uncertainty. An option written on the US dollar value of the Tokyo stock index is an example of such an option.

We now simulate a vector of underlying returns from the multivariate distribution

$$\left\{\hat{R}_{i,h}^K\right\}_{h=1}^{MC}, \quad \text{for } i = 1, 2, \ldots, n$$

and we calculate $\widehat{DV}$s by summing over the different assets using

$$\widehat{DV}_{PF,h}^K \approx \sum_{i=1}^{n} \delta_i S_{i,t} \hat{R}_{i,h}^K + \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{1}{2}\gamma_{ij} S_{i,t} S_{j,t} \hat{R}_{i,h}^K \hat{R}_{j,h}^K$$

The great benefit of this approach is that we are aggregating all the options on one particular asset into a delta and a gamma for that asset. Thus, if the portfolio consists of a thousand different types of option contracts, but only written on 100 different underlying assets, then the dimension of the approximated portfolio distribution is only 100.

As these formulas suggest, we could, in principle, simulate the distribution of the future asset returns at any horizon and calculate the portfolio Value-at-Risk for that horizon. However, a key problem with the delta and the delta-gamma approaches is that if we calculate the *VaR* for a horizon longer than one day, the delta and gamma numbers may not be reliable approximations to the risk of the option position because they are assumed to be constant through time when in reality they are not. We therefore next consider an approach that is computationally intensive, but does not suffer from the problems arising from approximating the options by delta and gamma.

6 Portfolio Risk Using Full Valuation

Linear and quadratic approximations to the nonlinearity arising from options can in some cases give a highly misleading picture of the risk from options. Particularly, if the portfolio contains options with different strike prices, then problems are likely to arise. We will give an explicit example of this type of problem.

In such complex portfolios, we may be forced to calculate the risk measure using what we will call full valuation. Full valuation consists of simulating future hypothetical underlying asset prices and using the option pricing model to calculate the corresponding future hypothetical option prices. For each hypothetical future asset price, every option written on that asset must be priced. While full valuation is precise, it is unfortunately also computationally intensive. Full valuation can, in principle, be done with any of the option pricing models discussed in Chapter 10.

6.1 The Single Underlying Asset Case

Consider first the simple case where our position consists of a short position in one call option. The dollar return at horizon K can be written

$$DV_{PF,t+K} = -1 \cdot \left(c(S_{t+K}, r_f, X, \tilde{T} - \tau; \sigma) - c^{mkt} \right)$$

where c^{mkt} is the current market price.

The τ is the risk horizon measured in calendar days because the option maturity, $\tilde{T}$, is measured in calendar days. The risk management horizon in trading days is denoted by K. For example, if we have a two-week \$*VaR* horizon, then K is 10 and τ is 14.

We can think of full valuation as pretending that we have arrived on the risk management horizon date and want to price all the options in the portfolio. As we do not know the price of the underlying asset K days into the future, we value the options for a range of hypothetical future prices of the underlying. Assuming a particular physical distribution of the return on the underlying asset, and applying the Monte Carlo

method discussed in Chapter 8, we can simulate future hypothetical returns on the underlying asset

$$\left\{\hat{R}_h^K\right\}_{h=1}^{MC}$$

and calculate future hypothetical asset prices

$$\left\{\hat{S}_h^K = S_t \exp\left(\hat{R}_h^K\right)\right\}_{h=1}^{MC}$$

We can now calculate the hypothetical changes in the portfolio value as

$$\widehat{DV}_{PF,h}^K = -1 \cdot \left(c(\hat{S}_h^K, r_f, X, \tilde{T} - \tau; q; \sigma) - c^{mkt}\right), \quad \text{for } h = 1, 2, \ldots, MC$$

The $\$VaR$ can now be calculated as in Chapter 8 using

$$\$VaR_{t+1:t+K}^p = -Percentile\left\{\left\{\widehat{DV}_{PF,h}^K\right\}_{h=1}^{MC}, 100p\right\}$$

Thus, we sort the portfolio value changes in $\left\{\widehat{DV}_{PF,h}^K\right\}_{h=1}^{MC}$ in ascending order and choose the $\$VaR_{t+1:t+K}^p$ to be the number such that only $p \cdot 100\%$ of the observations are smaller than the $\$VaR_{t+1:t+K}^p$.

6.2 The General Case

More generally, consider again the portfolio of linear assets such as stocks. We have

$$DV_{PF,t+K} = \sum_{i=1}^{n} \tilde{w}_i \left(S_{i,t+K} - S_{i,t}\right)$$

where $\tilde{w}_i$ is the number of asset i units held.

If we add, for example, call options to the portfolio, we would have

$$DV_{PF,t+K} = \sum_{i=1}^{n} \tilde{w}_i \left(S_{i,t+K} - S_{i,t}\right) + \sum_{i=1}^{n} \sum_{j} m_{i,j} \left(c(S_{i,t+K}, r_f, X_{i,j}, \tilde{T}_{i,j} - \tau, q_{i,j}; \sigma_i) - c_{i,j}^{mkt}\right)$$

where $m_{i,j}$ is the number of options of type j on the underlying asset i.

The Value-at-Risk from full valuation can be calculated from simulation again. Using the model for the returns distribution, we can simulate future returns and thus future asset prices

$$\left\{\hat{S}_{i,h}^K\right\}_{h=1}^{MC}, \quad \text{for } i = 1, 2, \ldots, n$$

and calculate the hypothetical changes in the portfolio value as

$$\widehat{DV}^K_{PF,h} = \sum_{i=1}^{n} \tilde{w}_i \left(\widehat{S}_{i,h} - S_{i,t}\right) + \sum_{i=1}^{n} \sum_{j} m_{i,j} \left(c(\hat{S}^K_{i,h}, r_f, X_{i,j}, \tilde{T}_{i,j} - \tau; q_{i,j}; \sigma_i) - c^{mkt}_{i,j} \right)$$

From these simulated value changes, we can calculate the dollar Value-at-Risk as

$$\$VaR^p_{t+1:t+K} = -Percentile\left\{ \left\{ \widehat{DV}^K_{PF,h} \right\}_{h=1}^{MC}, 100p \right\}$$

The full valuation approach has the benefit of being conceptually very simple; furthermore it does not rely on approximations to the option price. It does, however, require much more computational effort as all the future hypothetical prices of every option contract have to be calculated for every simulated future underlying asset price. Considerations of computational speed therefore sometimes dictate the choice between the more precise but slow full valuation method and the approximation methods, which are faster to implement.

7 A Simple Example

To illustrate the three approaches to option risk management, consider the following example. On January 6, 2010, we want to compute the 10-day $\$VaR$ of a portfolio consisting of a short position in one S&P 500 call option. The option has 43 calendar days to maturity, and it has a strike price of \$1135. The price of the option is \$26.54, and the underlying index is \$1137.14. The expected flow of dividends per day is 0.005697%, and the risk-free interest rate is 0.000682% per day. For simplicity, we assume a constant standard deviation of 1.5% per calendar day (for option pricing and delta calculation) or equivalently $0.015 \cdot \sqrt{365/252} = 0.0181$ per trading day (for calculating VaR in trading days). We will use the BSM model for calculating δ, γ as well as the full valuation option prices. We thus have

$$S_t = 1137.14$$
$$X = 1135$$
$$\tilde{T} = 43$$
$$r_f = 0.000682\%$$
$$q = 0.005697\%$$
$$\sigma = 1.5\% \text{ per calendar day}$$

from which we can calculate the delta and gamma of the option as

$$d = \frac{\ln(S_t/X) + \tilde{T}\left(r_f - f - q + \sigma^2/2\right)}{\sigma\sqrt{\tilde{T}}} = 0.046411$$

$$\delta = \exp(-q\tilde{T})\Phi(d) = 0.517240$$
$$\gamma = \frac{\phi(d)\,e^{-q\tilde{T}}}{S_t\sigma\sqrt{\tilde{T}}} = 0.003554$$

for the portfolio, which is short one option; we thus have

$$m \cdot \delta = -1 \cdot 0.517240$$
$$m \cdot \gamma = -1 \cdot 0.003554$$

where the -1 comes from the position being short.

In the delta-based model, the dollar *VaR* is

$$\begin{aligned}\$VaR^{0.01}_{t+1:t+10} &= -abs(\delta)S_t\sigma\sqrt{10}\Phi^{-1}_{0.01} \\ &\approx -abs(-1\cdot 0.517240)\cdot 1137.14\cdot 0.0181\cdot\sqrt{10}\cdot(-2.33) \\ &\approx \$77.92\end{aligned}$$

where we now use volatility in trading days.

Using the quadratic model and relying on the Cornish-Fisher approximation to the portfolio dollar return distribution, we calculate the first three moments for the K-day change in portfolio values, when the underlying return follows the $N(0, K\sigma^2)$ distribution. Setting $K = 10$, we get

$$\mu_{DV} \approx \frac{1}{2}S_t^2\gamma K\sigma^2 = -7.488792$$
$$\sigma^2_{DV} \approx S_t^2\delta^2 K\sigma^2 + \frac{1}{2}S_t^4\gamma^2K^2\sigma^4 = 1239.587604$$
$$\zeta_{1,DV} \approx \frac{\frac{9}{2}S_t^4\delta^2\gamma K^2\sigma^4 + \frac{15}{8}S_t^6\gamma^3K^3\sigma^6 - 3\left(S_t^2\delta^2K\sigma^2 + \frac{3}{4}S_t^4\gamma^2K^2\sigma^4\right)\mu_{DV} + 2\mu^3_{DV}}{\sigma^3_{DV}}$$
$$= -1.237724$$

where we use volatility denoted in trading days since K is denoted in trading days. The dollar *VaR* is then

$$\$VaR^{.01}_{t+1:t+10} = -\mu_{DV} - \left(\Phi_p^{-1} + \frac{1}{6}\left(\left(\Phi_p^{-1}\right)^2 - 1\right)\zeta_{1,DV}\right)\sigma_{DV} = \$121.44$$

which is **much** higher than the *VaR* from the linear model. The negative skewness coming from the option γ and captured by $\zeta_{1,DV}$ increases the quadratic *VaR* in comparison with the linear *VaR*, which implicitly assumes a skewness of 0.

Using instead the simulated quadratic model, we generate 5000 10-trading day returns, $\hat{R}_h, h = 1, 2, \ldots, 5000$ with a standard deviation of $0.0181 \cdot \sqrt{10}$. Using the

δ and γ calculated earlier, we find

$$\begin{aligned}
\$VaR^{0.01}_{t+1:t+10} &= -Percentile\left\{\left\{\delta S_t \hat{R}_h + \frac{1}{2}\gamma S_t^2 \cdot \hat{R}_h^2\right\}_{h=1}^{MC}, 1\right\} \\
&= -Percentile\left\{\left\{(-1 \cdot 0.517240) \cdot 1137.14 \cdot \hat{R}_h \right.\right. \\
&\qquad \left.\left. + \frac{1}{2}(-1 \cdot 0.003554) \cdot 1137.14^2 \cdot \hat{R}_h^2\right\}_{h=1}^{MC}, 1\right\} \\
&\approx \$126.06
\end{aligned}$$

Notice again that due to the relatively high γ of this option, the quadratic *VaR* is more than 50% higher than the linear *VaR*.

Finally, we can use the full valuation approach to find the most accurate *VaR*. Using the simulated asset returns $\hat{R}_h$ to calculate hypothetical future stock prices, $\hat{S}_h$, we calculate the simulated option portfolio value changes as

$$\widehat{DV}_{PF,h} = -1 \cdot \left(c(\hat{S}_h, r_f, X, \tilde{T} - 14, q; \sigma) - 35.10\right), \quad \text{for } h = 1, 2, \ldots, 5000$$

where 14 is the number of calendar days in the 10-trading-day risk horizon. We then calculate the full valuation *VaR* as

$$\begin{aligned}
\$VaR^{0.01}_{t+1:t+10} &= -Percentile\left\{\left\{\widehat{DV}_{PF,h}\right\}_{h=1}^{MC}, 1\right\} \\
&\approx \$144.83
\end{aligned}$$

In this example, the full valuation *VaR* is slightly higher than the quadratic *VaR*. The quadratic *VaR* thus provides a pretty good approximation in this simple portfolio of one option.

To gain further insight into the difference among the three *VaR*s, we plot the entire distribution of the hypothetical future 10-day portfolio dollar returns under the three models. Figure 11.6 shows a normal distribution with mean zero and variance $\delta^2 S_t^2 10\sigma^2 = 1121.91$.

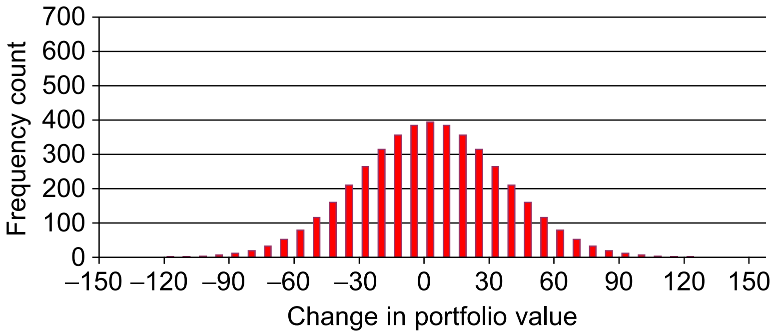
Figure 11.7 shows the histogram from the quadratic model using the 5000 simulated portfolio returns.

Finally, Figure 11.8 shows the histogram of the 5000 simulated full valuation dollar returns. Notice the stark differences between the delta-based method and the other two. The linear model assumes a normal distribution where there is no skewness. The quadratic model allows for skewness arising from the gamma of the option. The portfolio dollar return distribution has a negative skewness of around -1.29.

Finally, the full valuation distribution is slightly more skewed at -1.42. The difference in skewness arises from the asymmetry of the distribution now being simulated directly from the option returns rather than being approximated by the gamma of the option.

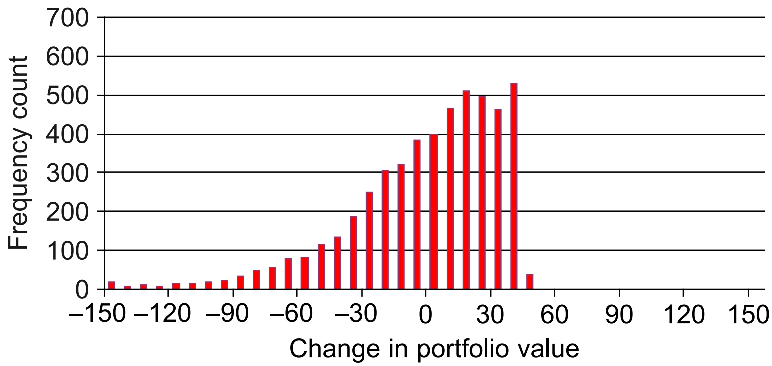
Further details on all the calculations in this section can be found on the web site.

Figure 11.6 Histogram of portfolio value changes using the delta-based model.



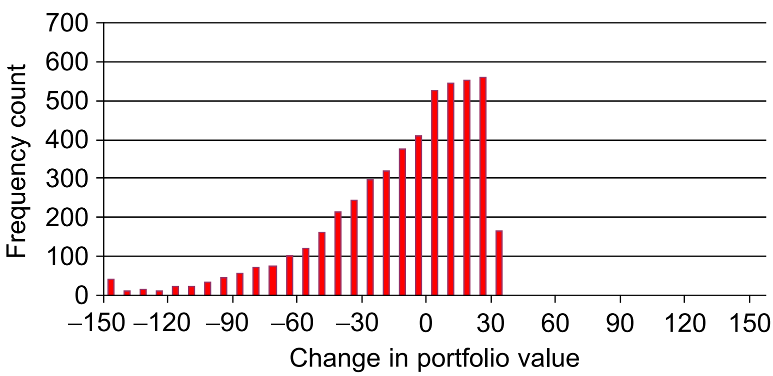
Notes: The plot shows the distribution of outcomes from 5,000 Monte Carlo replications of the delta-based model for a simple portfolio of a single call option.

Figure 11.7 Histogram of portfolio value changes using the gamma-based model.



Notes: The plot shows the distribution of outcomes from 5,000 Monte Carlo replications of the gamma-based model for a simple portfolio of a single call option.

Figure 11.8 Histogram of portfolio value changes using full valuation.



Notes: The plot shows the distribution of outcomes from 5,000 Monte Carlo replications of the full valuation model for a simple portfolio of a single call option.

8 Pitfall in the Delta and Gamma Approaches

While the previous example suggests that the quadratic approximation yields a good approximation to the true option portfolio distribution, we now show a different example, which illustrates that even the gamma approximation can sometimes be highly misleading.

To illustrate the potential problem with the approximations, consider an option portfolio that consists of three types of options all on the same asset, and that has a price of $S_t = 100$, all with $\tilde{T} = 28$ calendar days to maturity. The risk-free rate is $0.02/365$ and the volatility is 0.015 per calendar day. We take a short position in 1 put with a strike of 95, a short position in 1.5 calls with a strike of 95, and a long position in 2.5 calls with a strike of 105. Using the BSM model to calculate the delta and gamma of the individual options, we get

Type of Option	**Put**	**Call**	**Call**
Strike, X_j	95	95	105
Option Price	1.1698	6.3155	1.3806
Delta, δ_j	–0.2403	0.7597	0.2892
Gamma, γ_j	0.03919	0.03919	0.04307
Position, m_j	–1	–1.5	2.5

We are now interested in assessing the accuracy of the delta and gamma approximation for the portfolio over a five trading day or equivalently seven calendar day horizon. Rather than computing *VaR*s, we will take a closer look at the complete payoff profile of the portfolio for different future values of the underlying asset price, S_{t+5}. We refer to the value of the portfolio today as VPF_t and to the hypothetical future value as $VPF_{t+5}(S_{t+5})$.

We first calculate the value of the portfolio today as

$$\begin{aligned} VPF_t &= -1 \cdot 1.1698 - 1.5 \cdot 6.3155 + 2.5 \cdot 1.3806 \\ &= -7.1916 \end{aligned}$$

The delta of the portfolio is similarly

$$\begin{aligned} \delta &= -1 \cdot (-0.2403) - 1.5 \cdot 0.7597 + 2.5 \cdot 0.2892 \\ &= -0.1761 \end{aligned}$$

Now, the delta approximation to the portfolio value in five trading days is easily calculated as

$$\begin{aligned} VPF_{t+5}(S_{t+5}) &\approx VPF_t + \delta(S_{t+5} - S_t) \\ &= -7.1916 - 0.1761(S_{t+5} - 100) \end{aligned}$$

The gamma of the portfolio is

$$\begin{aligned} \gamma &= -1 \cdot 0.03919 - 1.5 \cdot 0.03919 + 2.5 \cdot 0.04307 \\ &= 0.0096898 \end{aligned}$$

and the gamma approximation to the portfolio value in five trading days is

$$\begin{aligned} VPF_{t+5}(S_{t+5}) &= VPF_t + \delta(S_{t+5} - S_t) + \frac{1}{2}\gamma(S_{t+5} - S_t)^2 \\ &= -7.1916 - 0.1761 \cdot (S_{t+5} - 100) + 0.004845 \cdot (S_{t+5} - 100)^2 \end{aligned}$$

Finally, relying on full valuation, we must calculate the future hypothetical portfolio values as

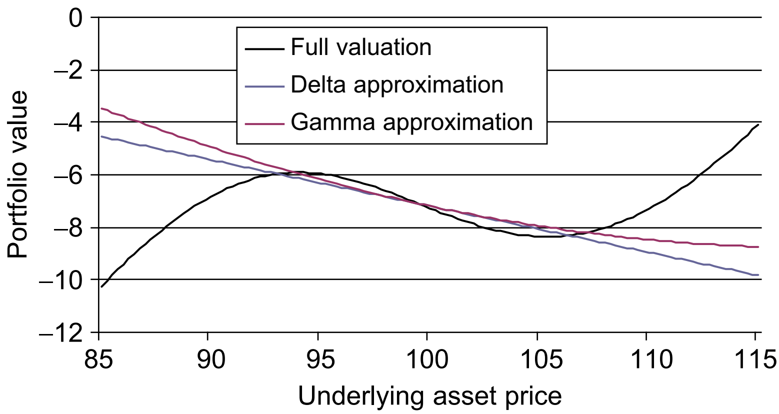
$$\begin{aligned} VPF_{t+5}(S_{t+5}) = & -1 \cdot p_{BSM}(S_{t+5}, r_f = 0.02/365, X = 95, \tilde{T} = 28 - 7; \sigma = 0.015) \\ & - 1.5 \cdot c_{BSM}(S_{t+5}, r_f = 0.02/365, X = 95, \tilde{T} = 28 - 7; \sigma = 0.015) \\ & + 2.5 \cdot c_{BSM}(S_{t+5}, r_f = 0.02/365, X = 105, \tilde{T} = 28 - 7; \sigma = 0.015) \end{aligned}$$

where we subtract seven calendar days from the time to maturity corresponding to the risk management horizon of five trading days.

Letting the hypothetical future underlying stock price vary from 85 to 115, the three-option portfolio values are shown in Figure 11.9. Notice how the exact portfolio value is akin to a third-order polynomial. The nonlinearity is arising from the fact that we have two strike prices. Both approximations are fairly poor when the stock price makes a large move, and the gamma-based model is even worse than the delta-based approximation when the stock price drops. Further details on all the calculations in this section can be found on the web site.

The important lesson of this three-option example is as follows: The different strike prices and the different exposures to the underlying asset price around the different strikes create higher order nonlinearities, which are not well captured by the simple

Figure 11.9 Future portfolio values for option portfolio.



Notes: We plot the portfolio value five days in the future for the three-option portfolio using the delta, gamma, and full valuation method plotted against the future price of the underlying asset.

linear and quadratic approximations. In realistic option portfolios consisting of thousands of contracts, there may be no alternative to using the full valuation method.

9 Summary

This chapter has presented three methods for incorporating options into the risk management model.

First, the delta-based approach consists of a complete linearization of the nonlinearity in the options. This crude approximation essentially allows us to use the methods in the previous chapters without modification. We just have to use the option's delta when calculating the portfolio weight of the option.

Second, we considered the quadratic, gamma-based approach, which attempts to capture the nonlinearity of the option while still mapping the option returns into the underlying asset returns. In general, we have to rely on simulation to calculate the portfolio distribution using the gamma approach, but we only simulate the underlying returns and not the option prices.

The third approach is referred to as full valuation. It avoids approximating the option price, but it involves much more computational work. We simulate returns on the underlying asset and then use an option pricing model to value each option in the portfolio for each of the future hypothetical underlying asset prices.

In a simple example of a portfolio consisting of just one short call option, we showed how a relatively large gamma would cause the delta-based *VaR* to differ substantially from the gamma and full valuation *VaR*s.

In another example involving a portfolio of three options with different strike prices and with large variations in the delta across the strike prices, we saw how the gamma and delta approaches were both quite misleading with respect to the future payoff profile of the options portfolio.

The main lesson from the chapter is that for nontrivial options portfolios and for risk management horizons beyond just a few days, the full valuation approach may be the only reliable choice.

Further Resources

The delta, gamma, and other risk measures are introduced and discussed in detail in Hull (2011). Backus et al. (1997) give the formula for delta in the Gram-Charlier model. Duan (1995) provided the delta in the GARCH option pricing model. Garcia and Renault (1998) discussed further the calculation of delta in the GARCH option pricing model. Heston and Nandi (2000) provided the formula for delta in the closed-form GARCH option pricing model.

Risk measurement in options portfolios has also been studied in Alexander et al. (2006), Sorwar and Dowd (2010), and Simonato (2011).

The sample portfolio used to illustrate the pitfalls in the use of the delta and gamma approximations is taken from Britten-Jones and Schaefer (1999), which also contains the analytical *VaR* formulas for the gamma approach assuming normality.

The important issue of accuracy versus computation speed in the full valuation versus delta and gamma approaches is analyzed in Pritsker (1997).

In this and the previous chapter, we have focused attention on European options. American options and many types of exotic options can be priced using binomial trees as we discussed in the last chapter. Deltas and gammas can be calculated from the tree approach as well as we have seen in this chapter. Hull (2011) contains a thorough introduction to binomial trees as an approximation to the normal distribution with constant variance. Ritchken and Trevor (1999) price American options under GARCH using a time-varying trinomial tree. Duan and Simonato (2001) priced American options under GARCH using instead a Markov chain simulation approach. Longstaff and Schwartz (2001) established a convenient least squares Monte Carlo method for pricing American and certain exotic options. See also Stentoft (2008) and Weber and Prokopczuk (2011) for methods to value American options in a GARCH framework.

Derman (1999), Derman and Kani (1994), and Rubinstein (1994) suggested binomial trees, which allow for implied volatility smiles as in the IVF approach in the previous chapter.

Whether one relies on the delta, gamma, or full valuation approach, an option pricing model is needed to measure the risk of the option position. Because all models are inevitably approximations, they introduce an extra source of risk referred to as model risk. Analysis of the various aspects of model risk can be found in Gibson (2000).

References

Alexander, S., Coleman, T., Li, Y., 2006. Minimizing CVaR and VaR for a portfolio of derivatives. J. Bank. Finance 30, 583–605.

Backus, D., Foresi, S., Li, K., Wu, L., 1997. Accounting for biases in Black-Scholes. Manuscript, The Stern School at NYU.

Britten-Jones, M., Schaefer, S., 1999. Non-linear value at risk. Eur. Finance Rev. 2, 161–187.

Derman, E., 1999. Regimes of volatility. Risk April, 12, 55–59.

Derman, E., Kani, I., 1994. The volatility smile and its implied tree. In: Quantitative Strategies Research Note. Goldman Sachs.

Duan, J., 1995. The GARCH option pricing model. Math. Finance 5, 13–32.

Duan, J., Simonato, J., 2001. American option pricing under GARCH by a Markov chain approximation. J. Econ. Dyn. Control 25, 1689–1718.

Garcia, R., Renault, E., 1998. A note on hedging in ARCH and stochastic volatility option pricing models. Math. Finance 8, 153–161.

Gibson, R., 2000. Model Risk: Concepts, Calibration and Pricing. Risk Books, London.

Heston, S., Nandi, S., 2000. A closed-form GARCH option pricing model. Rev. Financ. Stud. 13, 585–626.

Hull, J., 2011. Options, Futures and Other Derivatives, eighth ed. Prentice-Hall, Upper Saddle River, NJ.

Longstaff, F., Schwartz, E., 2001. Valuing American options by simulation: A simple least squares approach. Rev. Financ. Stud. 14, 113–147.

Pritsker, M., 1997. Evaluating value at risk methodologies: Accuracy versus computational time. J. Financ. Serv. Res. 12, 201–241.

Ritchken, P., Trevor, R., 1999. Pricing options under generalized GARCH and stochastic volatility processes. J. Finance 54, 377–402.
Rubinstein, M., 1994. Implied binomial trees. J. Finance 49, 771–818.
Simonato, J.-G., 2011. The performance of Johnson distributions for computing value at risk and expected shortfall. Available from: SSRN, http://ssrn.com/abstract=1706409.
Sorwar, G., Dowd, K., 2010. Estimating financial risk measures for options. J. Bank. Finance 34, 1982–1992.
Stentoft, L., 2008. American option pricing using GARCH models and the normal inverse Gaussian distribution. J. Financ. Econom. 6, 540–582.
Weber, M., Prokopczuk, M., 2011. American option valuation: Implied calibration of GARCH pricing models. J. Futures Mark. forthcoming.

Empirical Exercises

Open the Chapter11Data.xlsx file from the web site. The file contains European call options on the S&P 500 from January 6, 2010.

1. Assume a volatility of 0.015 per calendar day for option pricing and a volatility of $0.015 \cdot \sqrt{365/252} = 0.0181$ per trading day for return volatility. Calculate the delta and gamma of a short position of one option. Do this for every option in the sample. Calculate the delta-based portfolio variance for each option and the 10-trading-day (that is, 14-calendar-day) 1% delta-based dollar *VaR* for each option.
2. Assume a portfolio that consists of a short position of one in each of the option contracts. Calculate the 10-day, 1% dollar *VaR*s using the delta-based and the gamma-based models. Assume a normal distribution with the variance as in exercise 1. Use $MC = 5{,}000$ simulated returns for the 10-trading-day return. Compare the simulated quadratic *VaR* with the one using the Cornish-Fisher expansion formula.
3. Assume a short position of one option contract with 43 days to maturity and a strike price of 1135. Using the preceding 5,000 random normal numbers, calculate the changes in the 10-day portfolio value according to the delta-based, the gamma-based, and the full valuation approach. Calculate the 10-day, 1% dollar *VaR*s using the simulated data from the three approaches. Make histograms of the distributions of the changes in the portfolio value for these three approaches using the simulated data. Calculate the Cornish-Fisher *VaR* as well.
4. Replicate Figure 11.9.

The answers to these exercises can be found in the Chapter11Results.xlsx file on the companion site.

For more information see the companion site at
http://www.elsevierdirect.com/companions/9780123744487

12 Credit Risk Management

1 Chapter Overview

Credit risk can be defined as the risk of loss due to a counterparty's failure to honor an obligation in part or in full. Credit risk can take several forms. For banks credit risk arises fundamentally through its lending activities. Nonbank corporations that provide short-term credit to their debtors face credit risk as well. But credit risk can be important not only for banks and other credit providers; in certain cases it is important for investors as well.

Investors who hold a portfolio of corporate bonds or distressed equities need to get a handle on the default probability of the assets in their portfolio. Default risk, which is a key element of credit risk, introduces an important source of nonnormality into the portfolio in this case.

Credit risk can also arise in the form of counterparty default in a derivatives transaction. Figure 12.1 illustrates the counterparty risk exposure of a derivative contract. The horizontal axis shows the value of a hypothetical derivative contract in the absence of counterparty default risk. The derivative contract can have positive or negative value to us depending on the value of the asset underlying the derivative contract. The vertical axis shows our exposure to default of the derivative counterparty. When the derivative contract has positive value for us then counterparty risk is present. If the counterparty defaults then we might lose the entire value of the derivative and so the slope of the exposure is $+1$ to the right of zero in Figure 12.1. When the derivative contract has negative value to us then the default of our counterparty will have no effect on us; we will owe the same amount regardless. The counterparty risk is zero in this case.

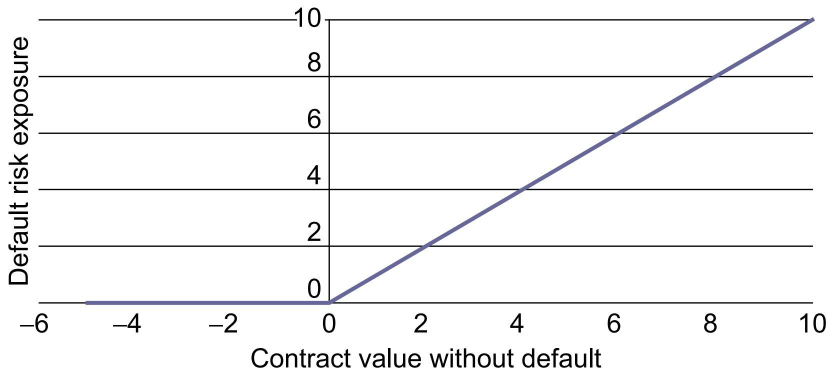
It is clear from Figure 12.1 that counterparty default risk has an option-like structure. Note that unlike in Chapters 10 and 11, Figure 12.1 has loss on the vertical axis instead of gain. Counterparty default risk can therefore be viewed as having sold a call option. Figure 12.1 ignores the fact that part of the value of the derivative position may be recovered in case of counterparty default. Partial recovery will be discussed later.

The chapter is structured as follows:

- Section 2 provides a few stylized facts on corporate defaults.
- Section 3 develops a model for understanding the effect on corporate debt and equity values of corporate default. Not surprisingly, default risk will have an

Elements of Financial Risk Management. DOI: 10.1016/B978-0-12-374448-7.00012-9

Figure 12.1 Exposure to counterparty default risk.



Notes: The figure shows the counterparty default risk exposure of a contract that can have positive or negative value without default risk. Default risk is only present when the contract has positive value.

important effect on how corporate debt (think a corporate bond) is priced, but default risk will also impact the equity price. The model will help us understand which factors drive corporate default risk.

- Section 4 builds on the single-firm model from Section 3 to develop a portfolio model of default risk. The model and its extensions provide a framework for computing credit Value-at-Risk.
- Section 5 discusses a range of further issues in credit risk including recovery rates, measuring credit quality through ratings, and measuring default risk using credit default swaps.

2 A Brief History of Corporate Defaults

Credit rating agencies such as Moody's and Standard & Poor's maintain databases of corporate defaults through time. In Moody's definition corporate default is triggered by one of three events: (1) a missed or delayed interest or principal payment, (2) a bankruptcy filing, or (3) a distressed exchange where old debt is exchanged for new debt that represents a smaller obligation for the borrower.

Table 12.1 shows some of the largest defaults in recent history. In terms of nominal size, the Lehman default in September 2008 clearly dominates. It is also interesting to note the apparent industry concentration of large defaults. The defaults in financial services companies is of course particularly concerning from the point of view of counterparty risk in derivatives transactions.

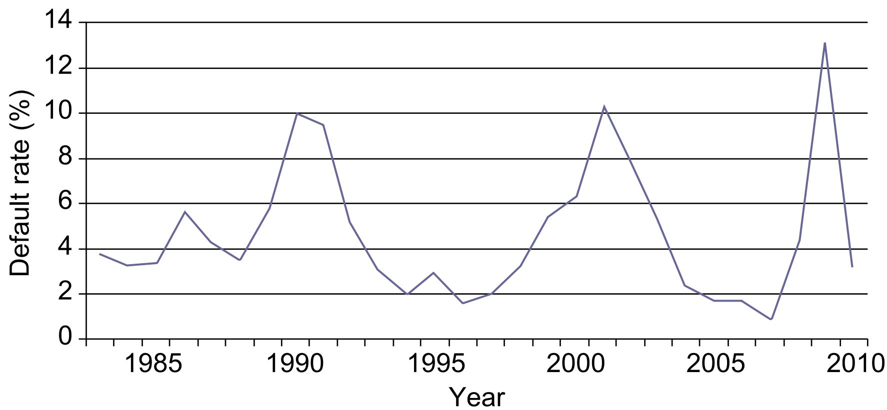
Figure 12.2 shows the global default rate for a one-year horizon plotted from 1983 through 2010 as compiled by Moody's. The figure includes only firms that are judged by Moody's to be low credit quality (also know as speculative grade) prior to default.

Table 12.1 Largest Moody's-rated defaults

Company	Default Volume ($ mill)	Year	Industry	Country
Lehman Brothers	$120,483	2008	Financials	United States
Worldcom, Inc.	$33,608	2002	Telecom/Media	United States
GMAC LLC	$29,821	2008	Financials	United States
Kaupthing Bank Hf	$20,063	2008	Financials	Iceland
Washington Mutual, Inc.	$19,346	2008	Financials	United States
Glitnir Banki Hf	$18,773	2008	Financials	Iceland
NTL Communications	$16,429	2002	Telecom/Media	United Kingdom
Adelphia Communications	$16,256	2002	Telecom/Media	United States
Enron Corp.	$13,852	2001	Energy	United States
Tribune Economy	$12,674	2008	Telecom/Media	United States

Notes: The table shows the largest defaults (by default volume) of the firms rated by Moody's during the 1920 through 2008 period. The data is from Moody's (2011).

Figure 12.2 Annual average corporate default rates for speculative grade firms, 1983–2010.



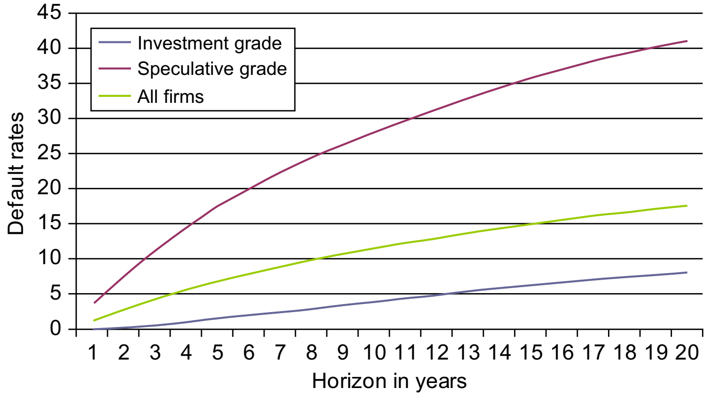
Notes: The figure shows the annual default rates for firms rated speculative grade by Moody's. The data is from Moody's (2011).

The overall default rate, which includes firms of high credit quality (also known as investment grade), is of course lower.

Figure 12.2 shows that the default rate is not constant over time—it is quite variable and seems highly persistent. The default rate alternates between periods of low default rates and periods with large spikes as happened for example during the recent financial crisis in 2007 to 2009.

The average corporate default rate for speculative grade (to be defined later) firms was 2.78% per year during the entire 1920–2010 period for which Moody's has data. For investment grade firms the average was just 0.15% per year.

Figure 12.3 Average cumulative global default rates.



Notes: The figure shows the cumulative (over the horizon in years) average global default rates for investment grade, speculative grade, and all firms. The rates are calculated using data from 1920 to 2010. The data is from Moody's (2011).

Figure 12.3 shows the cumulative default rates over a 1- through 20-year horizon plotted for investment grade, speculative grade, and all firms.

For example, the five-year cumulative default rate for all firms is 7.2%, which means historically there was a 7.2% probability of a firm defaulting during a five-year period. Over a 20-year horizon there is an 8.4% probability of an investment grade firm to default but a 41.4% probability of a speculative grade firm to default. The overall default rate is close to 18% for a 20-year horizon. The cumulative default rates are computed using data from 1920 through 2010.

These empirical default rates are not only of historical interest—they can also serve as useful estimates of the parameters in the credit portfolio risk models we study later.

3 Modeling Corporate Default

This section introduces the Merton model of corporate default, which provides important insights into the valuation of equity and debt when the probability of default is nontrivial. The model also helps us understand which factors affect the default probability.

Consider the situation where we are exposed to the risk that a particular firm defaults. This risk could arise from the fact that we own stock in the firm, or it could be that we have lent the firm cash, or it could be because the firm is a counterparty in a derivative transaction with us. We would like to use observed stock price on the firm to assess the probability of the firm defaulting.

Assume that the balance sheet of the company in question is of a particularly simple form. The firm is financed with debt and equity and all the debt expires at time $t+T$.

The face value of the debt is D and it is fixed. The future asset value of the firm, A_{t+T}, is uncertain.

3.1 Equity Is a Call Option on the Assets of the Firm

At time $t+T$ when the company's debt comes due the firm will continue to operate if $A_{t+T} > D$ but the firm's debt holders will declare the firm bankrupt if $A_{t+T} < D$ and the firm will go into default. The stock holders of the firm are the residual claimants on the firm and to the stock holders the firm is therefore worth

$$E_{t+T} = \max\{A_{t+T} - D, 0\}$$

when the debt comes due. As we saw in Chapter 10 this is exactly the payoff function of a call option with strike D that matures on day $t+T$.

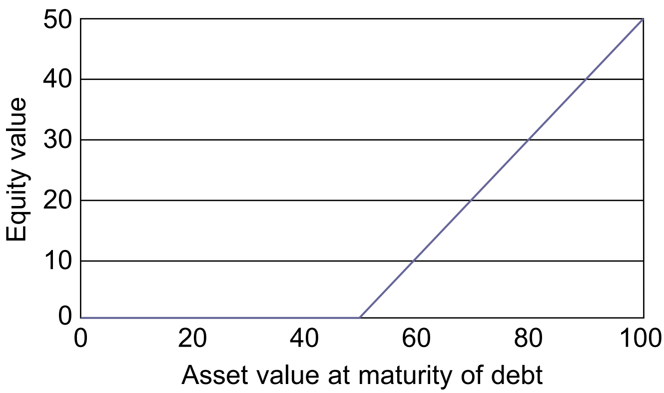
Figure 12.4 shows the value of firm equity E_{t+T} as a function of the asset value A_{t+T} at maturity of the debt when the face value of debt D is \$50.

The equity holder of a company can therefore be viewed as holding a call option on the asset value of the firm. It is important to note that in the case of stock options in Chapter 10 the stock price was the risky variable. In the present model of default, the asset value of the firm is the risky variable but the risky equity value can be derived as an option on the risky asset value.

The BSM formula in Chapter 10 can be used to value the equity in the firm in the Merton model. Assuming that asset volatility, σ_A, and the risk-free rate, r_f, are constant, and assuming that the log asset value is normally distributed we get the current value of the equity to be

$$E_t = A_t \Phi(d) - D\exp\left(-r_f T\right)\Phi\left(d - \sigma_A\sqrt{T}\right)$$

Figure 12.4 Equity value as function of asset value when face value of debt is \$50.



Notes: The figure plots the equity value of a firm as a function of its asset value at the time of the maturity of the firm's debt. The firm has outstanding debt with a face value of \$50.

where

$$d = \frac{\ln(A_t/D) + \left(r_f + \sigma_A^2/2\right)T}{\sigma_A\sqrt{T}}$$

Note that the risk-free rate, r_f, is not the rate earned on the company's debt; it is instead the rate earned on risk-free debt that can be obtained from the price of a government bond.

Recall from Chapters 10 and 11 that investors who are long options are long volatility. The Merton model therefore provides the additional insight that equity holders are long asset volatility. The option value is particularly large when the option is at-the-money; that is, when the asset value is close to the face value of debt. In this case if the manager holds equity he or she has an incentive to increase the asset value volatility (perhaps by taking on more risky projects) so as to increase the option value of equity. This action is not in the interest of the debt holders as we shall see now.

3.2 Corporate Debt Is a Put Option Sold

The simple accounting identity states that the asset value must equal the sum of debt and equity at any point in time and so we have

$$\begin{aligned} A_{t+T} &= D_{t+T} + E_{t+T} \\ &= D_{t+T} + \max\{A_{t+T} - D, 0\} \end{aligned}$$

where we have used the option payoff on equity described earlier. We use D_{t+T} to denote the market value of the debt at time $t+T$. Solving for the value of company debt, we get

$$D_{t+T} = A_{t+T} - \max\{A_{t+T} - D, 0\}$$

Figure 12.5 shows the payoff to the debt holder of the firm as a function of the asset value A_{t+T} when the face value of debt D is \$50.

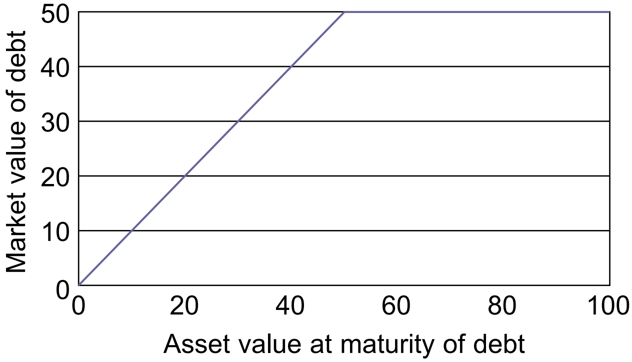
Comparing Figure 12.5 with the option payoffs in Chapter 10 we see that the debt holders look as if they have sold a put option although the out-of-the-money payoff has been lifted from 0 to \$50 on the vertical axis corresponding to the face value of debt in this example.

Figure 12.5 suggests that we can rewrite the debt holder payoff as

$$\begin{aligned} D_{t+T} &= A_{t+T} - \max\{A_{t+T} - D, 0\} \\ &= D - \max\{D - A_{t+T}, 0\} \end{aligned}$$

which shows that the holder of company debt can be viewed as being long a risk-free (think government) bond with face value D and short a put option on the asset value of the company, A_{t+T}, with a strike value of D. We can therefore use the model to value corporate debt; for example, corporate bonds.

Figure 12.5 Market value of debt as a function of asset value when face value of debt is $50.



Notes: The figure plots the market value of debt of a firm as a function of its asset value at the time of the maturity of the debt. The firm has outstanding debt with a face value of $50.

Using the put option formula from Chapter 10 the value today of the corporate debt with face value D is

$$D_t = e^{-r_f T}D - \left[e^{-r_f T}D\Phi\left(\sigma_A\sqrt{T} - d\right) - A_t\Phi\left(-d\right)\right]$$
$$= e^{-r_f T}D\Phi\left(d - \sigma_A\sqrt{T}\right) - A_t\Phi\left(-d\right)$$

where d is again defined by

$$d = \frac{\ln\left(A_t/D\right) + \left(r_f + \sigma_A^2/2\right)T}{\sigma_A\sqrt{T}}$$

The debt holder is short a put option and so is short asset volatility. If the manager takes actions that increase the asset volatility of the firm, then the debt holders suffer because the put option becomes more valuable.

3.3 Implementing the Model

Recall from Chapter 10 that the stock return volatility needed to be estimated for the BSM model to be implemented. In order to implement the Merton model we need values for σ_A and A_t, which are not directly observable. In practice, if the stock of the firm is publicly traded then we do observe the number of shares outstanding and we also observe the stock price, and we therefore do observe E_t,

$$E_t = S_t N_S$$

where N_S is the number of shares outstanding. From the call option relationship earlier we know that E_t is related to σ_A and A_t via the equation

$$E_t = A_t\Phi\left(d\right) - D\exp\left(-r_f T\right)\Phi\left(d - \sigma_A\sqrt{T}\right)$$

This gives us one equation in two unknowns. We need another equation. The preceding equation for E_t implies a dynamic for the stock price that can be used to derive the following relationship between the equity and asset volatilities:

$$\sigma E_t = \Phi(d)\,\sigma_A A_t$$

where σ is the stock price volatility as in Chapters 10 and 11. The stock price volatility can be estimated from historical returns or implied from stock option prices. We therefore now have two equations in two unknowns, A_t and σ_A. The two equations are nonlinear and so must be solved numerically using, for example, Solver in Excel.

Note that a crucially powerful feature of the Merton model is that we can use it to price corporate debt on firms even without observing the asset value as long as the stock price is available.

3.4 The Risk-Neutral Probability of Default

The risk-neutral probability of default in the Merton model corresponds in Chapter 10 to the probability that the put option is exercised. It is simply

$$\Pr(A_{t+T} < D) = 1 - \Phi\left(d - \sigma_A\sqrt{T}\right) = \Phi\left(\sigma_A\sqrt{T} - d\right)$$

It is important to note that this probability of default is constructed from the risk-neutral distribution (where assets grow at the risk-free rate r_f) of asset values and so it may well be different from the actual (unobserved) physical probability. The physical default probability could be derived in the model but would require an estimate of the physical growth rate of firm assets.

Default risk is also sometimes measured in terms of distance to default, which is defined as

$$dd = d - \sigma_A\sqrt{T} = \frac{\ln(A_t/D) + \left(r_f - \sigma_A^2/2\right)T}{\sigma_A\sqrt{T}}$$

The interpretation of dd is that it is the number of standard deviations the asset value must move down for the firm to default. As expected, the distance to default is increasing in the asset value and decreasing in the face value of debt. The distance to default is also decreasing in the asset volatility. Note that the probability of default is

$$\Pr(A_{t+T} < D) = \Phi(-dd)$$

The probability of default is therefore increasing in asset volatility.

4 Portfolio Credit Risk

The Merton model gives powerful intuition about corporate default and debt pricing and it enables us to link the debt value to equity price and volatility, which in the case

of public companies can be observed or estimated. While much can be learned from the Merton model, we have several motivations for going further.

- First, we are interested in studying the portfolio implications of credit risk. Default is a highly nonlinear event and furthermore default is correlated across firms and so credit risk is likely to impose limits on the benefits to diversification.
- Second, certain credit derivatives, such as collateralized debt obligations (CDOs), depend on the correlation of defaults that we therefore need to model.
- Third, for privately held companies we may not have the information necessary to implement the Merton model.
- Fourth, even if we have the information needed, for a portfolio of many loans, the implementation of Merton's model for each loan would be cumbersome.

In order to keep things relatively simple we will assume a single factor model similar to the market index model discussed in Chapter 7. For simplicity, we will also assume the normal distribution acknowledging that for risk management the use of the normal distribution is questionable at best.

We will assume a multivariate version of Merton's model in which the asset value of firm i is log normally distributed

$$\ln\left(A_{i,t+T}\right) = \ln A_{i,t} + r_f T - \tfrac{1}{2}\sigma_{A,i}^2 T + \sigma_{A,i}\sqrt{T} z_{i,t+T}$$

where $z_{i,t+T}$ is a standard normal variable. As before, the probability of default for firm i is

$$\Pr\left(A_{i,t+T} < D_i\right) = \Pr\left(\ln\left(A_{i,t+T}\right) < \ln\left(D_i\right)\right) = \Phi\left(-dd_i\right)$$

where

$$dd_i = \frac{\ln\left(A_{i,t}/D_i\right) + \left(r_f - \sigma_{A,i}^2/2\right)T}{\sigma_{A,i}\sqrt{T}}$$

We will assume further that the unconditional probability of default on any one loan is PD. This implies that the distance to default is now

$$\Pr\left(A_{i,t+T} < D_i\right) = PD \Longleftrightarrow$$
$$dd_i = -\Phi^{-1}\left(PD\right)$$

for all firms. A firm defaults when the asset value shock z_i is less than $-dd_i$ or equivalently less than $\Phi^{-1}\left(PD\right)$.

We will assume that the horizon of interest, T, is one year so that $T = 1$ and T is therefore left out of the formulas in this section. For ease of notation we will also suppress the time subscripts, t, in the following.

4.1 Factor Structure

The relationship between asset values across firms will be crucial for measuring portfolio credit risk. Assume that the correlation between any firm i and any other firm j is ρ, which does not depend on i nor j. This equi-correlation assumption implies a factor structure on the n asset values. We have

$$z_i = \sqrt{\rho}F + \sqrt{1-\rho}\tilde{z}_i$$

where the common (unobserved) factor F and the idiosyncratic $\tilde{z}_i$ are independent standard normal variables. Note that the z_is will be correlated with each other with coefficient ρ because they are all correlated with the common factor F with coefficient ρ.

Using the factor structure we can solve for $\tilde{z}_i$ in terms of z_i and F as

$$\tilde{z}_i = \frac{z_i - \sqrt{\rho}F}{\sqrt{1-\rho}}$$

From this we know that a firm defaults when $z_i < \Phi^{-1}(PD)$ or equivalently when $\tilde{z}_i$ is less than

$$\frac{\Phi^{-1}(PD) - \sqrt{\rho}F}{\sqrt{1-\rho}}$$

The probability of firm i defaulting conditional on the common factor F is therefore

$$\Pr[\text{Firm } i \text{ default}|F] = \Phi\left(\frac{\Phi^{-1}(PD) - \sqrt{\rho}F}{\sqrt{1-\rho}}\right)$$

Note that because of the assumptions we have made this probability is the same for all firms.

4.2 The Portfolio Loss Rate Distribution

Define the gross loss (before recovery) when firm i defaults to be L_i. Assuming that all loans are of equal size, relative loan size will not matter, and we can simply assume that

$$L_i = \begin{cases} 1, & \text{when firm } i \text{ defaults} \\ 0, & \text{otherwise} \end{cases}$$

The average size of the loans will be included in the following model.

The credit portfolio loss rate is defined as the average of the individual losses via

$$L = \frac{1}{n}\sum_{i=1}^{n} L_i$$

Note that L takes a value between zero and one.

The factor structure assumed earlier implies that conditional on F the L_i variables are independent. This allows the distribution of the portfolio loss rate to be derived. It is only possible to derive the exact distribution when assuming that the number of firms, n, is infinite. The distribution is therefore only likely to be accurate in portfolios of many relatively small loans.

As the number of loans goes to infinity we can derive the limiting CDF of the loss rate L to be

$$F_L(x; PD, \rho) = \Pr[L < x] = \Phi\left(\frac{\sqrt{1-\rho}\Phi^{-1}(x) - \Phi^{-1}(PD)}{\sqrt{\rho}}\right)$$

where $\Phi^{-1}(\bullet)$ is the standard normal inverse CDF as in previous chapters. The portfolio loss rate distribution thus appears to have similarities with the normal distribution but the presence of the $\Phi^{-1}(x)$ term makes the distribution highly nonnormal. This distribution is sometimes known as the Vasicek distribution, from Oldrich Vasicek who derived it.

The corresponding PDF of the portfolio loss rate is

$$f_L(x; PD, \rho) = \sqrt{\frac{1-\rho}{\rho}} \exp\left(-\frac{1}{2\rho}\left(\sqrt{1-\rho}\Phi^{-1}(x) - \Phi^{-1}(PD)\right)^2 + \frac{1}{2}\left(\Phi^{-1}(x)\right)^2\right)$$

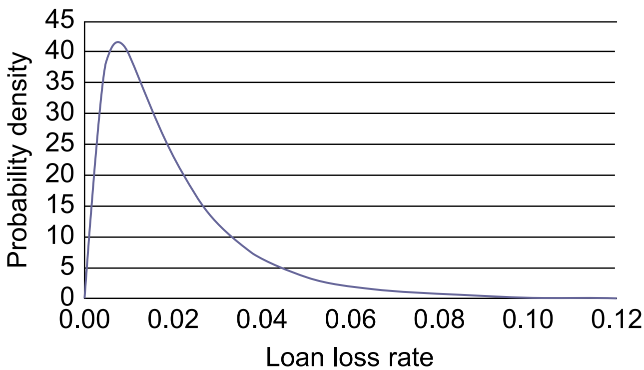
The mean of the distribution is PD and the variance is

$$\sigma_L^2 = \Phi_\rho\left(\Phi^{-1}(PD), \Phi^{-1}(PD)\right) - PD^2$$

where $\Phi_\rho(\bullet)$ is the CDF of the bivariate normal distribution we used in Chapters 7–9.

Figure 12.6 shows the PDF of the loan loss rate when $\rho = 0.10$ and $PD = 0.02$.

Figure 12.6 Portfolio loss rate distribution for $PD = 0.02$ and $\rho = 0.1$.



Notes: We plot the probability density function of the portfolio loss rate when the probability of default is 0.02 and the asset value correlation is 0.1.

Figure 12.6 clearly illustrates the nonnormality of the credit portfolio loss rate distribution. The loss distribution has large positive skewness that risk-averse investors will dislike. Recall from previous chapters that investors dislike negative skewness in the return distribution—equivalently they dislike positive skewness in the loss distribution.

The credit portfolio distribution in Figure 12.6 is not only important for credit risk measurement but it can also be used to value credit derivatives with multiple underlying assets such as collateralized debt obligations (CDOs), which were popular until the recent financial crisis.

4.3 Value-at-Risk on Portfolio Loss Rate

The *VaR* for the portfolio loss rate can be computed by inverting the CDF of the portfolio loss rate defined as $F_L(x)$ earlier. Because we are now modeling losses and not returns, we are looking for a loss *VaR* with probability $(1-p)$, which corresponds to a return *VaR* with probability p used in previous chapters.

We need to solve for VaR^{1-p} in

$$\Phi\left(\frac{\sqrt{1-\rho}\,\Phi^{-1}\left(VaR^{1-p}\right)-\Phi^{-1}(PD)}{\sqrt{\rho}}\right)=1-p$$

which yields the following *VaR* formula:

$$VaR^{1-p}=\Phi\left(\frac{\sqrt{\rho}\,\Phi^{-1}(1-p)+\Phi^{-1}(PD)}{\sqrt{1-\rho}}\right)$$

Figure 12.7 shows the $VaR_{0.99}$ as a function of ρ for various values of PD. Recall from the definitions of L_i and L that the loan loss rate can be at most 1 and so the *VaR* is restricted to be between 0 and 1 in this model.

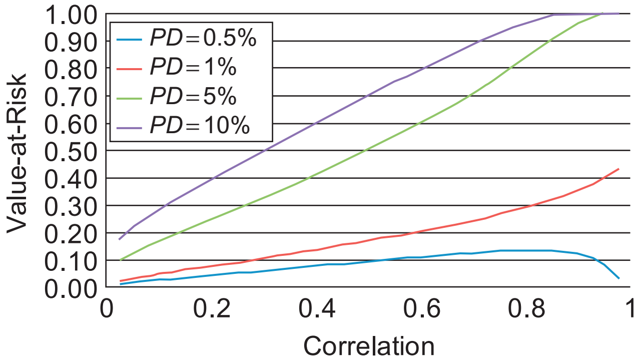
Not surprisingly, Figure 12.7 shows that the *VaR* is higher when the default probability PD is higher. The effects of asset value correlation on *VaR* are more intriguing. Generally, an increasing correlation implies loss of diversification and so does an increasing *VaR*, but when PD is low (0.5%) and the correlation is high, then a further increase in correlation actually decreases the *VaR* slightly. The effect of correlation on the distribution is clearly nonlinear.

4.4 Granularity Adjustment

The model may appear to be restrictive because we have assumed that the n loans are of equal size. But it is possible to show that the limiting loss rate distribution is the same even when the loans are not of the same same size as long as the portfolio is not dominated by a few very large loans.

The limiting distribution, which assumes that n is infinite, can of course only be an approximation in any real-life applications where n is finite. It is possible to derive finite-sample refinements to the limiting distribution based on so-called granularity

Figure 12.7 Loss rate *VaR* as a function of correlation with various default probabilities.



Notes: We plot the portfolio loss rate *VaR* against the asset value correlation for four different levels of unconditional default probabilities.

adjustments. Granularity adjustments for the *VaR* can be derived as well. For the particular model earlier, the granularity adjustment for the *VaR* is

$$GA(1-p) = \frac{1}{2}\left[\frac{\sqrt{\frac{1-\rho}{\rho}}\Phi^{-1}(1-p) - \Phi^{-1}\left(VaR^{1-p}\right)}{\phi\left(\Phi^{-1}\left(VaR^{1-p}\right)\right)} VaR^{1-p}\left(1 - VaR^{1-p}\right) + 2VaR^{1-p} - 1\right]$$

The granularity adjusted *VaR* can now be computed as

$$GAVaR^{1-p} = VaR^{1-p} + \frac{1}{n}GA(1-p)$$

where

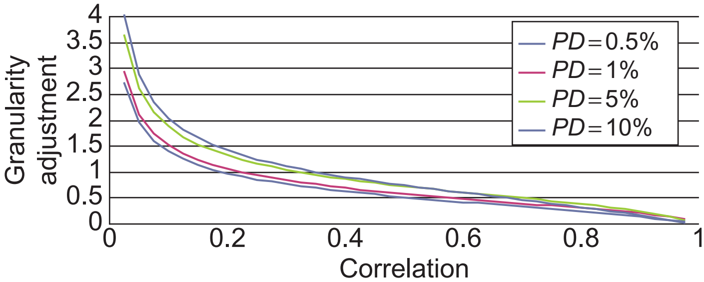
$$VaR^{1-p} = \Phi\left(\frac{\sqrt{\rho}\Phi^{-1}(1-p) + \Phi^{-1}(PD)}{\sqrt{1-\rho}}\right)$$

as before. Note that as n goes to infinity the granularity adjustment term $\frac{1}{n}GA(1-p)$ goes to zero.

Figure 12.8 shows the granularity adjustment $GA(1-p)$ corresponding to the *VaR*s in Figure 12.7.

It is clear from Figure 12.8 that the granularity adjustment is positive so that if we use the *VaR* that assumes infinitely many loans then we will tend to underestimate the true *VaR*. Figure 12.8 shows that the granularity adjustment is the largest when ρ is the smallest. The figure also shows that the adjustment is largest when the default probability, *PD*, is the largest.

Figure 12.8 Granularity adjustment to the loss rate *VaR* as a function of correlation.



Notes: The figure shows the granularity adjustment, *GA*, as a function of the asset value correlation. Four different unconditional default probability levels are shown. The loss rate *VaR* has a coverage of 99%. The *GA* has not been divided by *n*.

5 Other Aspects of Credit Risk

5.1 Recovery Rates

In the case of default part of the face value is typically recovered by the creditors. When assessing credit risk it is therefore important to have an estimate of the expected recovery rate. Moody's defines recovery using the market prices of the debt 30 days after the date of default. The recovery rate is then computed as the ratio of the post-default market price to the face value of the debt.

Figure 12.9 shows the average (issuer-weighted) recovery rates for investment grade, speculative grade, and all defaults. The figure is constructed for senior-unsecured debt only. The two-year recovery rate for speculative grade firms is 36.8%. This number shows the average recovery rate on speculative grade issues that default at some time within a two-year period. The defaults used in Figure 12.9 are from the 1982 to 2010 period.

Figure 12.9 shows that the average recovery rate is around 40% for senior unsecured debt. Sometimes the loss given default (*LGD*) is reported instead of the recovery rate (*RR*). They are related via the simple identity $LGD = 1 - RR$.

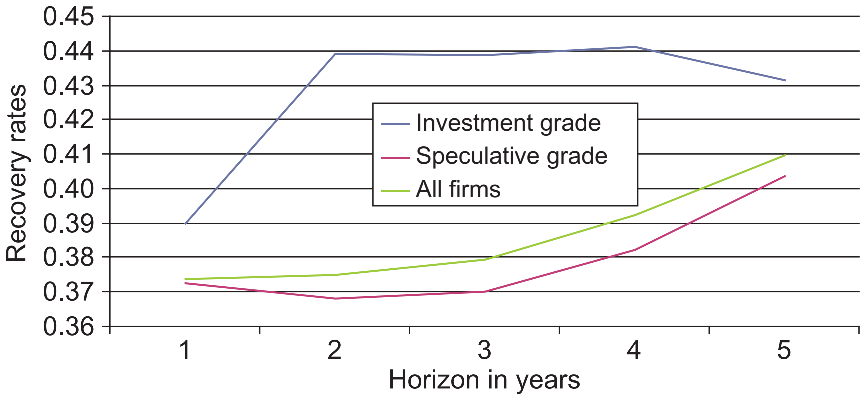
If we are willing to assume a constant proportional default rate across loans then we can compute the credit portfolio $\$VaR$ taking into account recovery rates using

$$
\begin{aligned}
\$VaR^{1-p} &= DV_{PF} \cdot LGD \cdot VaR^{1-p} \\
&= DV_{PF} \cdot (1 - RR) \cdot VaR^{1-p}
\end{aligned}
$$

where *RR* denotes the recovery rate and DV_{PF} denotes the dollar value of the portfolio as in Chapter 11. The VaR^{1-p} is the *VaR* from the portfolio loss rate defined earlier. The granularity adjusted dollar *VaR* can similarly be computed as

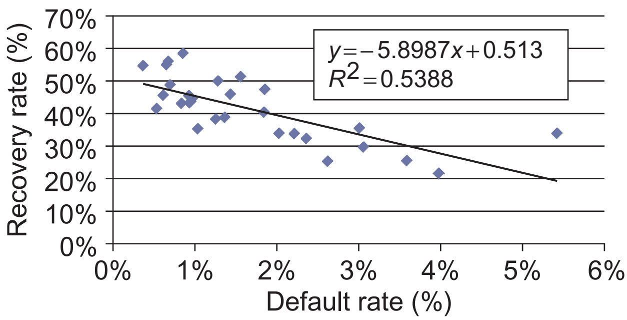
$$
\$GAVaR^{1-p} = DV_{PF} \cdot (1 - RR) \cdot GAVaR^{1-p}
$$

Figure 12.9 Average recovery rates for senior unsecured bonds.



Notes: The figure shows the average recovery rates on senior unsecured debt for investment grade, speculative grade, and all firms. The recovery rates are computed for firms that defaulted within one through five years. The rates are estimated by Moody's on data from 1982 through 2010.

Figure 12.10 Recovery rates versus default rates. Annual data, 1982–2010.



Notes: We scatter plot the average annual recovery rate against the average default rate on all bonds rated by Moody's during the 1982 to 2010 period. The line shows the best regression fit of y (the recovery rate) on x (the default rate).

Figure 12.2 showed that the default rate varied strongly with time. This has not been built into our static factor model for portfolio credit risk but it could be by allowing for the factor F to change over time.

We may also wonder if the recovery rate over time is correlated with default rate. Figure 12.10 shows that it is negatively correlated in the data. The higher the default rate the lower the recovery rate.

Stochastic recovery and its correlation with default present additional sources of credit risk that have not been included in our simple model but that could be built in. In

more complicated models, the *VaR* can be computed only via Monte Carlo simulation techniques such as those developed in Chapter 8.

5.2 Credit Quality Dynamics

The credit quality of a company typically declines well before any eventual default. It is important to capture this change in credit quality because the lower the quality the higher the chance of subsequent default.

One way to quantify the change in credit quality is by using the credit ratings provided by agencies such as Standard & Poor's and Moody's. The following list shows the ratings scale used by Moody's for long-term debt ordered from highest to lowest credit quality.

Investment Grades

- Aaa: Judged to be of the highest quality, with minimal credit risk.
- Aa: Judged to be of high quality and are subject to very low credit risk.
- A: Considered upper-medium grade and are subject to low credit risk.
- Baa: Subject to moderate credit risk. They are considered medium grade and as such may possess certain speculative characteristics.

Speculative Grades

- Ba: Judged to have speculative elements and are subject to substantial credit risk.
- B: Considered speculative and are subject to high credit risk.
- Caa: Judged to be of poor standing and are subject to very high credit risk.
- Ca: Highly speculative and are likely in, or very near, default, with some prospect of recovery of principal and interest.
- C: The lowest rated class of bonds and are typically in default, with little prospect for recovery of principal or interest.

Table 12.2 shows a matrix of one-year transition rates between Moody's ratings. The number in each cell corresponds to the probability that a company will transition from the row rating to the column rating in a year. For example, the probability of a Baa rated firm getting downgraded to Ba is 4.112%.

The transition probabilities are estimated from the actual ratings changes during each year in the 1970–2010 period. The Ca_C category combines the companies rated Ca and C. The second-to-last column, which is labeled WR, denotes companies for which the rating was withdrawn. More interestingly, the right-most column shows the probability of a firm with a particular row rating defaulting within a year. Notice that the default rates are monotonically increasing as the credit ratings worsen. The probability of an Aaa rated firm defaulting is virtually zero whereas the probability of a Ca or C rated firm defaulting within a year is 35.451%.

Table 12.2 Average one-year rating transition rates, 1970–2010

From/To	Aaa	Aa	A	Baa	Ba	B	Caa	Ca_C	WR	Default
Aaa	87.395%	8.626%	0.602%	0.010%	0.027%	0.002%	0.002%	0.000%	3.336%	0.000%
Aa	0.971%	85.616%	7.966%	0.359%	0.045%	0.018%	0.008%	0.001%	4.996%	0.020%
A	0.062%	2.689%	86.763%	5.271%	0.488%	0.109%	0.032%	0.004%	4.528%	0.054%
Baa	0.043%	0.184%	4.525%	84.517%	4.112%	0.775%	0.173%	0.019%	5.475%	0.176%
Ba	0.008%	0.056%	0.370%	5.644%	75.759%	7.239%	0.533%	0.080%	9.208%	1.104%
B	0.010%	0.034%	0.126%	0.338%	4.762%	73.524%	5.767%	0.665%	10.544%	4.230%
Caa	0.000%	0.021%	0.021%	0.142%	0.463%	8.263%	60.088%	4.104%	12.176%	14.721%
Ca_C	0.000%	0.000%	0.000%	0.000%	0.324%	2.374%	8.880%	36.270%	16.701%	35.451%

Notes: The table shows Moody's credit rating transition rates estimated on annual data from 1970 through 2010. Each row represents last year's rating and each column represents this year's rating. Ca_C combines two rating categories. WR refers to withdrawn rating. Data is from Moody's (2011).

Note also that for all the rating categories, the highest probability is to remain within the same category within the year. Ratings are thus persistent over time. The lowest rated firms have the least persistent rating. Many of them default or their ratings are withdrawn.

The rating transition matrices can be used in Monte Carlo simulations to generate a distribution of ratings outcomes for a credit risk portfolio. This can in turn be used along with ratings-based bond prices to compute the credit *VaR* of the portfolio. JP Morgan's CreditMetrics system for credit risk management is based on such an approach.

The credit risk portfolio model developed in Section 4 can also be extended to allow for debt value declines due to a deterioration of credit quality.

5.3 Credit Default Swaps

As discussed earlier, the price of corporate debt such as bonds is highly affected by the probability of default of the corporation issuing the debt. However, corporate bond prices are also affected by the prevailing risk-free interest rate as the Merton model showed us. Furthermore, in reality, corporate bonds are often relatively illiquid and so command a premium for illiquidity and illiquidity risk. Observed corporate bond prices therefore do not give a clean view of default risk. Fortunately, derivative contracts known as credit default swaps (CDS) that allow investors to trade default risk directly have appeared. CDS contracts give a pure market-based view of the default probability and its market price.

In a CDS contract the default protection buyer pays fixed quarterly cash payments (usually quoted in basis points per year) to the default protection seller. In return, in the event that the underlying corporation defaults, the protection seller will pay the protection buyer an amount equal to the par value of the underlying corporate bond minus the recovered value.

CDS contracts are typically quoted in spreads or premiums. A premium or spread of 200 basis points means that the protection buyer has to pay the protection seller 2% of the underlying face value of debt each year if a new CDS contract is entered into. Although the payments are made quarterly, the spreads are quoted in annual terms.

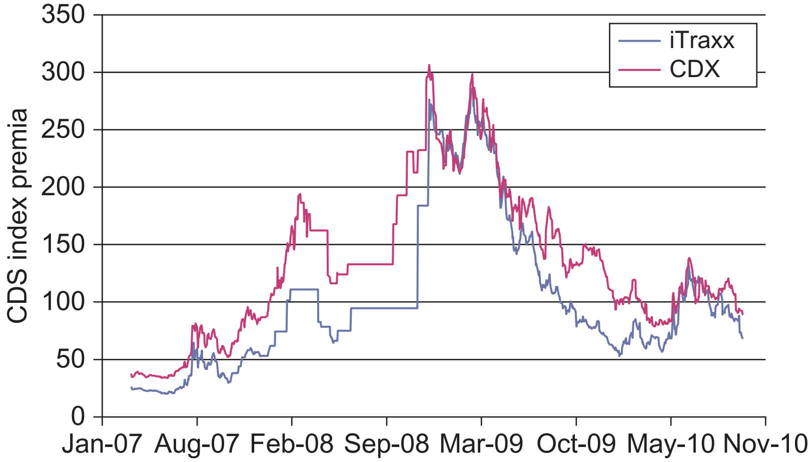
CDS contracts have become very popular because they allow the protection buyer to hedge default risk. They of course also allow investors to take speculative views on default risk as well as to arbitrage relative mispricings between corporate equity and bond prices.

Paralleling the developments in equity markets, CDS index contracts have developed as well. They allow investors to trade a basket of default risks using one liquid index contract rather than many illiquid firm-specific contracts.

Figure 12.11 shows the time series of CDS premia for two of the most prominent indices, namely the CDX NA (North American) and iTraxx EU (Europe). The CDX NA and iTraxx EU indexes each contain 125 underlying firms.

The data in Figure 12.11 are for five-year CDS contracts and the spreads are observed daily from March 20, 2007 through September 30, 2010. The financial

Figure 12.11 CDS index premia. CDX North America and iTraxx Europe.



Notes: The figure shows the daily CDS premia on the CDX North America and iTraxx Europe indices. The series are recorded from March 20, 2007 through September 30, 2010.

crisis is painfully evident in Figure 12.11. Perhaps remarkably, unlike CDOs, trading in CDSs remained strong throughout the crisis and CDS contracts have become an important tool for managing credit risk exposures.

6 Summary

Credit risk is of fundamental interest to lenders but it is also of interest to investors who face counterparty default risk or investors who hold portfolios with corporate bonds or distressed equities.

This chapter has provided some important stylized facts on corporate default and recovery rates. We also developed a theoretical framework for understanding default based on Merton's seminal model. Merton's model studies one firm in isolation but it can be generalized to a portfolio of firms using Vasicek's factor model structure. The factor model can be used to compute *VaR* in credit portfolios. The chapter also briefly discussed credit default swaps, which are the financial instruments most closely linked with default risk.

Further Resources

This chapter has presented only some of the basic ideas and models in credit risk analysis. Hull (2008) discusses in detail credit risk management in banks. See Lando (2004) for a thorough treatment of credit risk models.

Merton (1974) developed the single-firm model in Section 3 and Vasicek (2002) developed the factor model in Section 4. The default and recovery data throughout the chapter is from Moody's (2009, 2011).

Many researchers have developed extensions to the basic Merton model. Chapter 2 in Lando (2004) provides an excellent overview that includes Mason and Bhattacharya (1981) and Zhou (2001), who allow for jumps in the asset value. Hull et al. (2004) show that Merton's model can be extended so as to allow for the smirks in option-implied equity volatility that we saw in Chapter 10.

This chapter did not cover the so-called Z-scores from Altman (1968), who provides an important empirical approach to default prediction using financial ratios in a discriminant analysis.

The static and normal one-factor portfolio credit risk model in Section 4 can also be extended in several ways. Gagliardini and Gourieroux (2011a, b) survey alternatives that include adding factors, allowing for nonnormality, and allowing for dynamic factors to capture default dynamics. The granularity adjustments have been studied in Gordy (2003), Gordy and Lutkebohmert (2007), and Gourieroux and Monfort (2010), and are based on the sensitivity analysis of *VaR* in Gourieroux et al. (2000).

Gordy (2003) studies credit risk models from a regulatory perspective. Basel Committee on Banking Supervision (2001, 2003) contains the regulatory framework for ratings-based capital requirements.

Gordy (2000) provides a useful comparison of the two main credit portfolio risk models developed in industry, namely CreditRisk+ from Credit Suisse (1997) and CreditMetrics from JP Morgan (1997).

Jacobs and Li (2008) investigate corporate bond prices allowing for dynamic volatility of the form we studied in Chapter 4. The empirical relationship between asset correlation and default has been investigated by Lopez (2004). Karoui (2007) develops a model that allows for stochastic recovery rates. Christoffersen et al. (2009) study empirically the dynamic dependence of default.

References

Altman, E., 1968. Financial ratios: Discriminant analysis, and the prediction of corporate bankruptcy. J. Finance 23, 589–609.

Basel Committee on Banking Supervision, 2001. The New Basel Capital Accord, Consultative Document of the Bank for International Settlements, Part 2: Pillar 1. Available from: www.bis.org.

Basel Committee on Banking Supervision, 2003. The New Basel Capital Accord, Consultative Document of the Bank for International Settlements, Part 3: The Second Pillar. Available from: www.bis.org.

Christoffersen, P., Ericsson, J., Jacobs, K., Jin, X., 2009. Exploring Dynamic Default Dependence. Available from: SSRN, http://ssrn.com/abstract=1400427.

Credit Suisse, 1997. CreditRisk+: A Credit Risk Management Framework, Credit Risk Financial Products. Available from: www.csfb.com.

Gagliardini, P., Gourieroux, C., 2011a. Granularity Theory with Application to Finance and Insurance. Manuscript, University of Toronto.

Gagliardini, P., Gourieroux, C., 2011b. Granularity Adjustment for Risk Measures: Systematic vs. Unsystematic Risk. Manuscript, University of Toronto.

Gordy, M., 2000. A comparative anatomy of credit risk models. J. Bank. Finance 24, 119–149.

Gordy, M., 2003. A risk-factor model foundation for ratings-based bank capital rules. J. Financ. Intermed. 12, 199–232.
Gordy, M., Lutkebohmert, E., 2007. Granularity Adjustment for Basel II. Deutsche Bundesbank Discussion Paper.
Gourieroux, C., Laurent, J., Scaillet, O., 2000. Sensitivity analysis of values at risk. J. Empir. Finance 7, 225–245.
Gourieroux, C., Monfort, A., 2010. Granularity in a qualitative factor model. J. Credit Risk 5, Winter 2009/10, pp. 29–65.
Hull, J., 2008. Risk Management and Financial Institutions, second ed. Prentice Hall, Upper Saddle River, New Jersey.
Hull, J., Nelken, I., White, A., 2004. Merton's model, credit risk and volatility skews. J. Credit Risk 1, 1–27.
Jacobs, K., Li, X., 2008. Modeling the dynamics of credit spreads with stochastic volatility. Manag. Sci. 54, 1176–1188.
Karoui, L., 2007. Modeling defaultable securities with recovery risk. Available from: SSRN, http://ssrn.com/abstract=896026.
Lando, D., 2004. Credit Risk Modeling. Princeton University Press, Princeton, New Jersey.
Lopez, J., 2004. The empirical relationship between average asset correlation, firm probability of default and asset size. J. Financ. Intermed. 13, 265–283.
Mason, S., Bhattacharya, S., 1981. Risky debt, jump processes, and safety covenants. J. Financ. Econ. 9, 281–307.
Merton, R., 1974. On the pricing of corporate debt: The risk structure of interest rates. J. Finance 29, 449–470.
Moody's, 2009. Corporate Default and Recovery Rates, 1920–2008. Special Comment, Moody's. Available from: www.moodys.com.
Moody's, 2011. Corporate Default and Recovery Rates, 1920–2010, Special Comment, Moody's. Available from: www.moodys.com.
Morgan, J.P., 1997. CreditMetrics, Technical Document. Available from: http://www.msci.com/resources/technical_documentation/CMTD1.pdf.
Vasicek, O., 2002. Loan portfolio value. Risk December, pp. 160–162.
Zhou, C., 2001. The term structure of credit spreads with jump risk. J. Bank. Finance 25 (11), 2015–2040.

Empirical Exercises

Open the Chapter12Data.xlsx file from the web site.

1. Use the AR(1) model from Chapter 3 to model the default rate in Figure 12.2. What does the AR(1) model reveal regarding the persistence of the default rate? Can you fit the default rate closely? What is the fit (in terms of R^2) of the AR(1) model?
2. Consider a company with assets worth \$100 and a face value of debt of \$80. The log return on government debt is 3%. If the company debt expires in two years and the asset volatility is 80% per year, then what is the current market value of equity and debt?
3. Assume that a company has 100,000 shares outstanding and that the stock is trading at \$8 with a stock price volatility of 50% per year. The company has \$500,000 in debt (face value) that matures in six months. The risk-free rate is 3%. Solve for the asset value and the asset volatility of the firm. What is the distance to default and the probability of default? What is the market value of the debt?

4. Assume a portfolio with 100 small loans. The average default rate is 3% per year and the asset correlation of the firms underlying the loans is 5%. The recovery rate is 40%. Compute the \$*VaR* and the \$*GAVaR* for the portfolio.
5. Replicate Figure 12.6. How does the loss rate distribution change when the correlation changes?

The answers to these exercises can be found in the Chapter12Results.xlsx file on the companion site.

For more information see the companion site at
http://www.elsevierdirect.com/companions/9780123744487

13 Backtesting and Stress Testing

1 Chapter Overview

The first 12 chapters have covered various methods for constructing risk management models. Along the way we also considered several diagnostic checks. For example, in Chapter 1 we looked at the autocorrelations of returns to see if the assumption of a constant mean was valid. In Chapters 4 and 5 we looked at the autocorrelation function of returns squared divided by the time-varying variance to assess if we had modeled the variance dynamics properly. We also ran variance regressions to assess the forecasting performance of the suggested GARCH models. In Chapter 6, we studied the so-called QQ plots to see if the distribution we assumed for standardized returns captured the extreme observations in the sample. We also looked at the reaction of various risk models to an extreme event such as the 1987 stock market crash. In Chapter 9 we looked at Threshold Correlations. Finally, in Chapter 10 we illustrated option pricing model misspecification in terms of implied volatility smiles and smirks.

The objective in this chapter is to consider the ex ante risk measure forecasts from the model and compare them with the ex post realized portfolio return. The risk measure forecast could take the form of a Value-at-Risk (*VaR*), an Expected Shortfall (*ES*), the shape of the entire return distribution, or perhaps the shape of the left tail of the distribution only. We want to be able to backtest any of these risk measures of interest. The backtest procedures developed in this chapter can be seen as a final diagnostic check on the aggregate risk model, thus complementing the various specific diagnostics covered in previous chapters. The discussion on backtesting is followed up by a section on stress testing at the end of the chapter. The material in the chapter will be covered as follows:

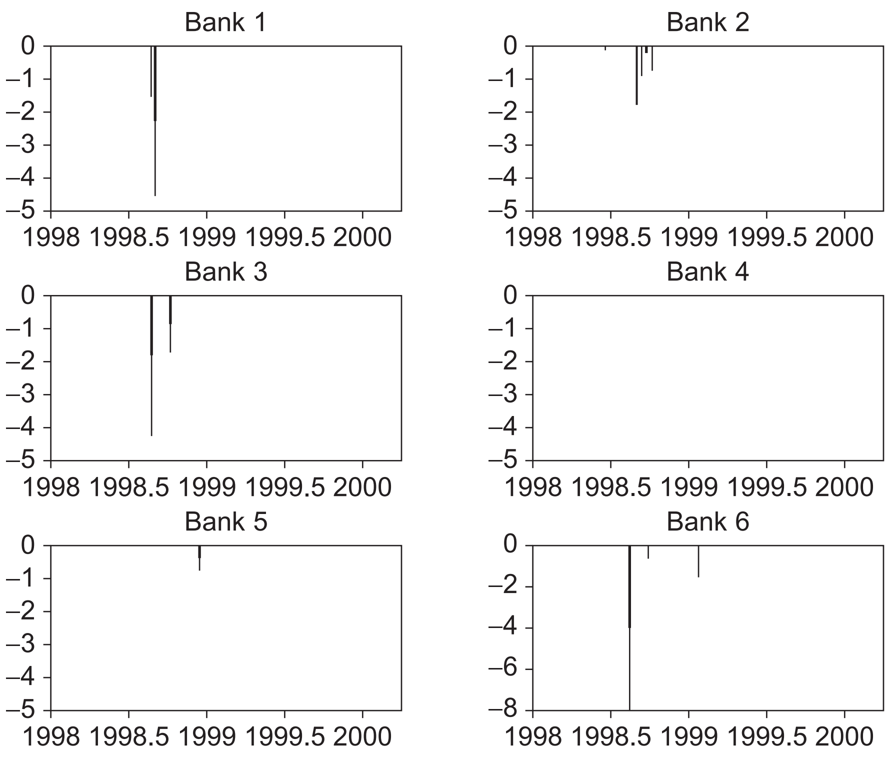
- We take a brief look at the performance of some real-life *VaR*s from six large (and anonymous) commercial banks. The clustering of *VaR* violations in these real-life *VaR*s provides sobering food for thought.
- We establish procedures for backtesting *VaR*s. We start by introducing a simple unconditional test for the average probability of a *VaR* violation. We then test the independence of the *VaR* violations. Finally, we combine the unconditional test and the independence test in a test of correct conditional *VaR* coverage.
- We consider using explanatory variables to backtest the *VaR*. This is done in a regression-based framework.

Elements of Financial Risk Management. DOI: 10.1016/B978-0-12-374448-7.00013-0

- We establish backtesting procedures for the Expected Shortfall measure.
- We broaden the focus to include the entire shape of the distribution of returns. The distributional forecasts can be backtested as well, and we suggest ways to do so. Risk managers typically care most about having a good forecast of the left tail of the distribution, and we therefore modify the distribution test to focus on backtesting the left tail of the distribution only.
- We define stress testing and give a critical survey of the way it is often implemented. Based on this critique we suggest a coherent framework for stress testing.

Before we get into the technical details of backtesting *VaR*s and other risk measures, it is instructive to take a look at the performance of some real-life *VaR*s. Figure 13.1 shows the exceedances (measured in return standard deviations) of the *VaR* in six large (and anonymous) U.S. commercial banks during the January 1998 to March 2001 period. Whenever the realized portfolio return is worse than the *VaR*, the difference between the two is shown. Whenever the return is better, zero is shown. The difference

Figure 13.1 Value-at-Risk exceedences from six major commercial banks.



Notes: The figure shows the *VaR* exceedences on the days where the loss was larger than the *VaR*. Each panel corresponds to a large U.S. commercial bank. The figure is reprinted from Berkowitz and O'Brien (2002).

is divided by the standard deviation of the portfolio across the period. The return is daily, and the *VaR* is reported for a 1% coverage rate. To be exact, we plot the time series of

$$Min\left\{R_{PF,t+1} - \left(-VaR^{.01}_{t+1}\right), 0\right\}/\sigma_{PF,t+1}$$

Bank 4 has no violations at all, and in general the banks have fewer violations than expected. Thus, the banks on average report a *VaR* that is higher than it should be. This could either be due to the banks deliberately wanting to be cautious or the *VaR* systems being biased. Another culprit is that the returns reported by the banks contain nontrading-related profits, which increase the average return without substantially increasing portfolio risk.

More important, notice the clustering of *VaR* violations. The violations for each of Banks 1, 2, 3, 5, and 6 fall within a very short time span and often on adjacent days. This clustering of *VaR* violations is a serious sign of risk model misspecification. These banks are most likely relying on a technique such as Historical Simulation (HS), which is very slow at updating the *VaR* when market volatility increases. This issue was discussed in the context of the 1987 stock market crash in Chapter 2.

Notice also how the *VaR* violations tend to be clustered across banks. Many violations appear to be related to the Russia default and Long Term Capital Management bailout in the fall of 1998. The clustering of violations across banks is important from a regulator perspective because it raises the possibility of a countrywide banking crisis.

Motivated by this sobering evidence of misspecification in existing commercial bank *VaR*s, we now introduce a set of statistical techniques for backtesting risk management models.

2 Backtesting *VaR*s

Recall that a VaR^p_{t+1} measure promises that the actual return will only be worse than the VaR^p_{t+1} forecast $p \cdot 100\%$ of the time. If we observe a time series of past ex ante *VaR* forecasts and past ex post returns, we can define the "hit sequence" of *VaR* violations as

$$I_{t+1} = \begin{cases} 1, \textit{if } R_{PF,t+1} < -VaR^p_{t+1} \\ 0, \textit{if } R_{PF,t+1} \geq -VaR^p_{t+1} \end{cases}$$

The hit sequence returns a 1 on day $t+1$ if the loss on that day was larger than the *VaR* number predicted in advance for that day. If the *VaR* was not violated, then the hit sequence returns a 0. When backtesting the risk model, we construct a sequence $\{I_{t+1}\}_{t=1}^{T}$ across T days indicating when the past violations occurred.

2.1 The Null Hypothesis

If we are using the perfect *VaR* model, then given all the information available to us at the time the *VaR* forecast is made, we should not be able to predict whether the *VaR*

will be violated. Our forecast of the probability of a *VaR* violation should be simply p every day. If we could predict the *VaR* violations, then that information could be used to construct a better risk model. In other words, the hit sequence of violations should be completely unpredictable and therefore distributed independently over time as a Bernoulli variable that takes the value 1 with probability p and the value 0 with probability $(1-p)$. We write

$$H_0 : I_{t+1} \sim \text{i.i.d. Bernoulli}(p)$$

If p is $1/2$, then the i.i.d. Bernoulli distribution describes the distribution of getting a "head" when tossing a fair coin. The Bernoulli distribution function is written

$$f(I_{t+1};p) = (1-p)^{1-I_{t+1}} p^{I_{t+1}}$$

When backtesting risk models, p will not be $1/2$ but instead on the order of 0.01 or 0.05 depending on the coverage rate of the *VaR*. The hit sequence from a correctly specified risk model should thus look like a sequence of random tosses of a coin, which comes up heads 1% or 5% of the time depending on the *VaR* coverage rate.

2.2 Unconditional Coverage Testing

We first want to test if the fraction of violations obtained for a particular risk model, call it π, is significantly different from the promised fraction, p. We call this the unconditional coverage hypothesis. To test it, we write the likelihood of an i.i.d. Bernoulli(π) hit sequence

$$L(\pi) = \prod_{t=1}^{T} (1-\pi)^{1-I_{t+1}} \pi^{I_{t+1}} = (1-\pi)^{T_0} \pi^{T_1}$$

where T_0 and T_1 are the number of 0s and 1s in the sample. We can easily estimate π from $\hat{\pi} = T_1/T$; that is, the observed fraction of violations in the sequence. Plugging the maximum likelihood (ML) estimates back into the likelihood function gives the optimized likelihood as

$$L\left(\hat{\pi}\right) = (1 - T_1/T)^{T_0} (T_1/T)^{T_1}$$

Under the unconditional coverage null hypothesis that $\pi = p$, where p is the known *VaR* coverage rate, we have the likelihood

$$L(p) = \prod_{t=1}^{T} (1-p)^{1-I_{t+1}} p^{I_{t+1}} = (1-p)^{T_0} p^{T_1}$$

We can check the unconditional coverage hypothesis using a likelihood ratio test

$$LR_{uc} = -2 \ln \left[L(p) / L\left(\hat{\pi}\right) \right]$$

Asymptotically, that is, as the number of observations, T, goes to infinity, the test will be distributed as a χ^2 with one degree of freedom. Substituting in the likelihood functions, we write

$$LR_{uc} = -2\ln\left[(1-p)^{T_0}p^{T_1}/\left\{(1-T_1/T)^{T_0}(T_1/T)^{T_1}\right\}\right] \sim \chi_1^2$$

The larger the LR_{uc} value is the more unlikely the null hypothesis is to be true. Choosing a significance level of say 10% for the test, we will have a critical value of 2.7055 from the χ_1^2 distribution. If the LR_{uc} test value is larger than 2.7055, then we reject the *VaR* model at the 10% level. Alternatively, we can calculate the P-value associated with our test statistic. The P-value is defined as the probability of getting a sample that conforms even less to the null hypothesis than the sample we actually got given that the null hypothesis is true. In this case, the P-value is calculated as

$$\text{P-value} \equiv 1 - F_{\chi_1^2}(LR_{uc})$$

where $F_{\chi_1^2}(\bullet)$ denotes the cumulative density function of a χ^2 variable with one degree of freedom. If the P-value is below the desired significance level, then we reject the null hypothesis. If we, for example, obtain a test value of 3.5, then the associated P-value is

$$\text{P-value} = 1 - F_{\chi_1^2}(3.5) = 1 - 0.9386 = 0.0614$$

If we have a significance level of 10%, then we would reject the null hypothesis, but if our significance level is only 5%, then we would not reject the null that the risk model is correct on average.

The choice of significance level comes down to an assessment of the costs of making two types of mistakes: We could reject a correct model (Type I error) or we could fail to reject (that is, accept) an incorrect model (Type II error). Increasing the significance level implies larger Type I errors but smaller Type II errors and vice versa. In academic work, a significance level of 1%, 5%, or 10% is typically used. In risk management, the Type II errors may be very costly so that a significance level of 10% may be appropriate.

Often, we do not have a large number of observations available for backtesting, and we certainly will typically not have a large number of violations, T_1, which are the informative observations. It is therefore often better to rely on Monte Carlo simulated P-values rather than those from the χ^2 distribution. The simulated P-values for a particular test value can be calculated by first generating 999 samples of random i.i.d. Bernoulli(p) variables, where the sample size equals the actual sample at hand. Given these artificial samples we can calculate 999 simulated test statistics, call them $\left\{\widetilde{LR}_{uc}(i)\right\}_{i=1}^{999}$. The simulated P-value is then calculated as the share of simulated LR_{uc} values that are larger than the actually obtained LR_{uc} test value. We can write

$$\text{P-value} = \frac{1}{1000}\left\{1 + \sum_{i=1}^{999}\mathbf{1}\left(\widetilde{LR}_{uc}(i) > LR_{uc}\right)\right\}$$

where $\mathbf{1}(\bullet)$ takes on the value of one if the argument is true and zero otherwise.

To calculate the tests in the first place, we need samples where *VaR* violations actually occurred; that is, we need some ones in the hit sequence. If we, for example, discard simulated samples with zero or one violations before proceeding with the test calculation, then we are in effect conditioning the test on having observed at least two violations.

2.3 Independence Testing

Imagine all the *VaR* violations or "hits" in a sample happening around the same time, which was the case in Figure 13.1. Would you then be happy with a *VaR* with correct average (or unconditional) coverage? The answer is clearly no. For example, if the 5% *VaR* gave exactly 5% violations but all of these violations came during a three-week period, then the risk of bankruptcy would be much higher than if the violations came scattered randomly through time. We therefore would very much like to reject *VaR* models that imply violations that are clustered in time. Such clustering can easily happen in a *VaR* constructed from the Historical Simulation method in Chapter 2, if the underlying portfolio return has a clustered variance, which is common in asset returns and which we studied in Chapter 4.

If the *VaR* violations are clustered, then the risk manager can essentially predict that if today is a violation, then tomorrow is more than $p \cdot 100\%$ likely to be a violation as well. This is clearly not satisfactory. In such a situation, the risk manager should increase the *VaR* in order to lower the conditional probability of a violation to the promised p.

Our task is to establish a test that will be able to reject a *VaR* with clustered violations. To this end, assume the hit sequence is dependent over time and that it can be described as a so-called first-order Markov sequence with transition probability matrix

$$\Pi_1 = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix}$$

These transition probabilities simply mean that conditional on today being a nonviolation (that is, $I_t = 0$), then the probability of tomorrow being a violation (that is, $I_{t+1} = 1$) is π_{01}. The probability of tomorrow being a violation given today is also a violation is defined by

$$\pi_{11} = \Pr\left(I_{t+1} = 1 | I_t = 1\right)$$

Similarly, the probability of tomorrow being a violation given today is not a violation is defined by

$$\pi_{01} = \Pr\left(I_{t+1} = 1 | I_t = 0\right)$$

The first-order Markov property refers to the assumption that only today's outcome matters for tomorrow's outcome—that the exact sequence of past hits does not matter, only the value of I_t matters. As only two outcomes are possible (zero and one), the two

probabilities π_{01} and π_{11} describe the entire process. The probability of a nonviolation following a nonviolation is $1-\pi_{01,}$ and the probability of a nonviolation following a violation is $1-\pi_{11}$.

If we observe a sample of T observations, then we can write the likelihood function of the first-order Markov process as

$$L(\Pi_1) = (1-\pi_{01})^{T_{00}}\pi_{01}^{T_{01}}(1-\pi_{11})^{T_{10}}\pi_{11}^{T_{11}}$$

where T_{ij}, $i,j=0,1$ is the number of observations with a j following an i. Taking first derivatives with respect to π_{01} and π_{11} and setting these derivatives to zero, we can solve for the maximum likelihood estimates

$$\hat{\pi}_{01} = \frac{T_{01}}{T_{00}+T_{01}}$$
$$\hat{\pi}_{11} = \frac{T_{11}}{T_{10}+T_{11}}$$

Using then the fact that the probabilities have to sum to one, we have

$$\hat{\pi}_{00} = 1-\hat{\pi}_{01}$$
$$\hat{\pi}_{10} = 1-\hat{\pi}_{11}$$

which gives the matrix of estimated transition probabilities

$$\hat{\Pi}_1 \equiv \begin{bmatrix} \hat{\pi}_{00} & \hat{\pi}_{01} \\ \hat{\pi}_{10} & \hat{\pi}_{11} \end{bmatrix} = \begin{bmatrix} 1-\hat{\pi}_{01} & \hat{\pi}_{01} \\ 1-\hat{\pi}_{11} & \hat{\pi}_{11} \end{bmatrix} = \begin{bmatrix} \frac{T_{00}}{T_{00}+T_{01}} & \frac{T_{01}}{T_{00}+T_{01}} \\ \frac{T_{10}}{T_{10}+T_{11}} & \frac{T_{11}}{T_{10}+T_{11}} \end{bmatrix}$$

Allowing for dependence in the hit sequence corresponds to allowing π_{01} to be different from π_{11}. We are typically worried about positive dependence, which amounts to the probability of a violation following a violation (π_{11}) being larger than the probability of a violation following a nonviolation (π_{01}). If, on the other hand, the hits are independent over time, then the probability of a violation tomorrow does not depend on today being a violation or not, and we write $\pi_{01} = \pi_{11} = \pi$. Under independence, the transition matrix is thus

$$\hat{\Pi} = \begin{bmatrix} 1-\hat{\pi} & \hat{\pi} \\ 1-\hat{\pi} & \hat{\pi} \end{bmatrix}$$

We can test the independence hypothesis that $\pi_{01} = \pi_{11}$ using a likelihood ratio test

$$LR_{ind} = -2\ln\left[L\left(\hat{\Pi}\right)/L\left(\hat{\Pi}_1\right)\right] \sim \chi_1^2$$

where $L\left(\hat{\Pi}\right)$ is the likelihood under the alternative hypothesis from the LR_{uc} test.

In large samples, the distribution of the LR_{ind} test statistic is also χ^2 with one degree of freedom. But we can calculate the P-value using simulation as we did before. We again generate 999 artificial samples of i.i.d. Bernoulli variables, calculate 999 artificial test statistics, and find the share of simulated test values that are larger than the actual test value.

As a practical matter, when implementing the LR_{ind} tests we may incur samples where $T_{11} = 0$. In this case, we simply calculate the likelihood function as

$$L\left(\hat{\Pi}_1\right) = \left(1 - \hat{\pi}_{01}\right)^{T_{00}} \hat{\pi}_{01}^{T_{01}}$$

2.4 Conditional Coverage Testing

Ultimately, we care about simultaneously testing if the *VaR* violations are independent and the average number of violations is correct. We can test jointly for independence and correct coverage using the conditional coverage test

$$LR_{cc} = -2\ln\left[L(p)/L\left(\hat{\Pi}_1\right)\right] \sim \chi_2^2$$

which corresponds to testing that $\pi_{01} = \pi_{11} = p$.

Notice that the LR_{cc} test takes the likelihood from the null hypothesis in the LR_{uc} test and combines it with the likelihood from the alternative hypothesis in the LR_{ind} test. Therefore,

$$\begin{aligned} LR_{cc} &= -2\ln\left[L(p)/L\left(\hat{\Pi}_1\right)\right] \\ &= -2\ln\left[\left\{L(p)/L\left(\hat{\Pi}\right)\right\}\left\{L(\hat{\pi})/L\left(\hat{\Pi}_1\right)\right\}\right] \\ &= -2\ln\left[L(p)/L\left(\hat{\Pi}\right)\right] - 2\ln\left[L\left(\hat{\Pi}\right)/L\left(\hat{\Pi}_1\right)\right] \\ &= LR_{uc} + LR_{ind} \end{aligned}$$

so that the joint test of conditional coverage can be calculated by simply summing the two individual tests for unconditional coverage and independence. As before, the P-value can be calculated from simulation.

2.5 Testing for Higher-Order Dependence

In Chapter 1 we used the autocorrelation function (ACF) to assess the dependence over time in returns and squared returns. We can of course use the ACF to assess dependence in the *VaR* hit sequence as well. Plotting the hit-sequence autocorrelations against their lag order will show if the risk model gives rise to autocorrelated hits, which it should not.

As in Chapter 3, the statistical significance of a set of autocorrelations can be formally tested using the Ljung-Box statistic. It tests the null hypothesis that the autocorrelation for lags 1 through m are all jointly zero via

$$LB(m) = T(T+2)\sum_{\tau=1}^{m} \frac{\widehat{\rho}_{\tau}^{2}}{T-\tau} \sim \chi_{m}^{2}$$

where $\widehat{\rho}_{\tau}$ is the autocorrelation of the *VaR* hit sequence for lag order τ. The chi-squared distribution with m degrees of freedom is denoted by χ_{m}^{2}. We reject the null hypothesis that the hit autocorrelations for lags 1 through m are jointly zero when the $LB(m)$ test value is larger than the critical value in the chi-squared distribution with m degrees of freedom.

3 Increasing the Information Set

The preceding tests are quick and easy to implement. But because they only use information on past *VaR* violations, they might not have much power to detect misspecified risk models. To increase the testing power, we consider using the information in past market variables, such as interest rate spreads or volatility measures. The basic idea is to test the model using information that may explain when violations occur. The advantage of increasing the information set is not only to increase the power of the tests but also to help us understand the areas in which the risk model is misspecified. This understanding is key in improving the risk models further.

If we define the vector of variables available to the risk manager at time t as X_t, then the null hypothesis of a correct risk model can be written as

$$H_0 : \Pr(I_{t+1} = 1|X_t) = p \Leftrightarrow E[I_{t+1}|X_t] = p$$

The first hypothesis says that the conditional probability of getting a *VaR* violation on day $t+1$ should be independent of any variable observed at time t, and it should simply be equal to the promised *VaR* coverage rate, p. This hypothesis is equivalent to the conditional expectation of a *VaR* violation being equal to p. The reason for the equivalence is that I_{t+1} can only take on one of two values: 0 and 1. Thus, we can write the conditional expectation as

$$E[I_{t+1}|X_t] = 1 \cdot \Pr(I_{t+1} = 1|X_t) + 0 \cdot \Pr(I_{t+1} = 0|X_t) = \Pr(I_{t+1} = 1|X_t)$$

Thinking of the null hypothesis in terms of a conditional expectation immediately leads us to consider a regression-based approach, because regressions are essentially conditional mean functions.

3.1 A Regression Approach

Consider regressing the hit sequence on the vector of known variables, X_t. In a simple linear regression, we would have

$$I_{t+1} = b_0 + b_1' X_t + e_{t+1}$$

where the error term e_{t+1} is assumed to be independent of the regressor, X_t.

The hypothesis that $E[I_{t+1}|X_t] = p$ is then equivalent to

$$E\left[b_0 + b_1' X_t + e_{t+1}|X_t\right] = p$$

As X_t is known, taking expectations yields

$$b_0 + b_1' X_t = p$$

which can only be true if $b_0 = p$ and b_1 is a vector of zeros. In this linear regression framework, the null hypothesis of a correct risk model would therefore correspond to the hypothesis

$$H_0 : b_0 = p, b_1 = 0$$

which can be tested using a standard F-test (see the econometrics textbooks referenced at the end of Chapter 3). The P-value from the test can be calculated using simulated samples as described earlier.

There is, of course, no particular reason why the explanatory variables should enter the conditional expectation in a linear fashion. But nonlinear functional forms could be tested as well.

4 Backtesting Expected Shortfall

In Chapter 2, we argued that the Value-at-Risk had certain drawbacks as a risk measure, and we defined Expected Shortfall (*ES*),

$$ES_{t+1}^p = -E_t\left[R_{PF,t+1}|R_{PF,t+1} < -VaR_{t+1}^p\right]$$

as a viable alternative. We now want to think about how to backtest the *ES* risk measure.

Consider again a vector of variables, X_t, which are known to the risk manager and which may help explain potential portfolio losses beyond what is explained by the risk model. The *ES* risk measure promises that whenever we violate the *VaR*, the expected value of the violation will be equal to ES_{t+1}^p. We can therefore test the *ES* measure by checking if the vector X_t has any ability to explain the deviation of the observed shortfall or loss, $-R_{PF,t+1}$, from the Expected Shortfall on the days where the *VaR* was violated. Mathematically, we can write

$$-R_{PF,t+1} - ES_{t+1}^p = b_0 + b_1' X_t + e_{t+1}, \quad \text{for } t+1 \text{ where } R_{PF,t+1} < -VaR_{t+1}^p$$

where $t+1$ now refers only to days where the *VaR* was violated. The observations where the *VaR* was not violated are simply removed from the sample. The error term e_{t+1} is again assumed to be independent of the regressor, X_t.

To test the null hypothesis that the risk model from which the *ES* forecasts were made uses all information optimally ($b_1 = 0$), and that it is not biased ($b_0 = 0$), we can jointly test that $b_0 = b_1 = 0$.

Notice that now the magnitude of the violation shows up on the left-hand side of the regression. But notice that we can still only use information in the tail to backtest. The *ES* measure does not reveal any particular properties about the remainder of the distribution, and therefore, we only use the observations where the losses were larger than the *VaR*.

5 Backtesting the Entire Distribution

Rather than focusing on particular risk measures from the return distribution such as the Value-at-Risk or the Expected Shortfall, we could instead decide to backtest the entire return distribution from the risk model. This would have the benefit of potentially increasing further the power to reject bad risk models. Notice, however, that we are again changing the object of interest: If only the *VaR* is reported, for example, from Historical Simulation, then we cannot test the distribution.

Consider a risk model that at the end of each day produces a cumulative distribution forecast for next day's return, call it $F_t(\bullet)$. Then at the end of every day, after having observed the actual portfolio return, we can calculate the risk model's probability of observing a return below the actual. We will denote this so-called transform probability by $\tilde{p}_{t+1}$:

$$\tilde{p}_{t+1} \equiv F_t\left(R_{PF,t+1}\right)$$

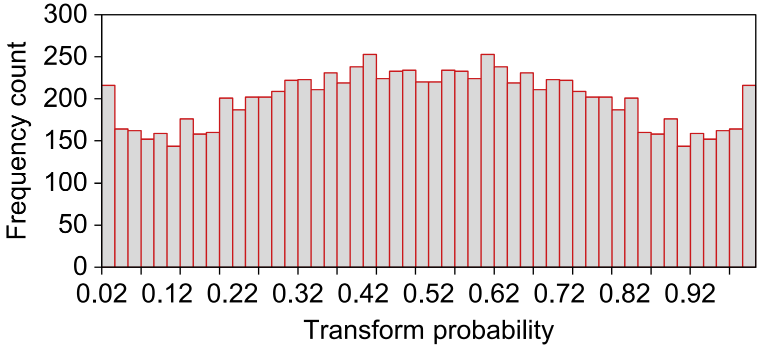
If we are using the correct risk model to forecast the return distribution, then we should not be able to forecast the risk model's probability of falling below the actual return. In other words, the time series of observed probabilities $\tilde{p}_{t+1}$ should be distributed independently over time as a Uniform(0,1) variable. We therefore want to consider tests of the null hypothesis

$$H_0 : \tilde{p}_{t+1} \sim \text{i.i.d. Uniform}\,(0, 1)$$

The Uniform(0,1) distribution function is flat on the interval 0 to 1 and zero everywhere else. As the $\tilde{p}_{t+1}$ variable is a probability, it must lie in the zero to one interval. A visual diagnostic on the distribution would be to simply construct a histogram and check to see if it looks reasonably flat. If systematic deviations from a flat line appear in the histogram, then we would conclude that the distribution from the risk model is misspecified.

For example, if the true portfolio return data follows a fat-tailed Student's $t(d)$ distribution, but the risk manager uses a normal distribution model, then we will see too

Figure 13.2 Histogram of the transform probability.



Notes: We plot the histogram of the transform probability when the returns follow an i.i.d. Student's $t(d)$ distribution with $d = 6$, but they are forecasted by an i.i.d. normal distribution.

many $\tilde{p}_{t+1}$s close to zero and one, too many around 0.5, and too few elsewhere. This would just be another way of saying that the observed returns data have more observations in the tails and around zero than the normal distribution allows for. Figure 13.2 shows the histogram of a $\tilde{p}_{t+1}$ sequence, obtained from taking $F_t\left(R_{PF,t+1}\right)$ to be normally distributed with zero mean and variance $d/(d-2)$, when it should have been Student's $t(d)$, with $d = 6$. Thus, we use the correct mean and variance to forecast the returns, but the shape of our density forecast is incorrect.

The histogram check is of course not a proper statistical test, and it does not test the time variation in $\tilde{p}_{t+1}$. If we can predict $\tilde{p}_{t+1}$ using information available on day t, then $\tilde{p}_{t+1}$ is not i.i.d., and the conditional distribution forecast, $F_t\left(R_{PF,t+1}\right)$, is therefore not correctly specified either. We want to consider proper statistical tests here.

Unfortunately, testing the i.i.d. uniform distribution hypothesis is cumbersome due to the restricted support of the uniform distribution. We therefore transform the i.i.d. Uniform $\tilde{p}_{t+1}$ to an i.i.d. standard normal variable $\tilde{z}_{t+1}$ using the inverse cumulative distribution function, Φ^{-1}. We write

$$H_0 : \tilde{p}_{t+1} \sim \text{i.i.d. Uniform}\,(0, 1) \Leftrightarrow$$
$$H_0 : \tilde{z}_{t+1} = \Phi^{-1}\left(\tilde{p}_{t+1}\right) = \Phi^{-1}\left(F_t\left(R_{PF,t+1}\right)\right) \sim \text{i.i.d. } N\,(0, 1)$$

We are now left with a test of a variable conforming to the standard normal distribution, which can easily be implemented.

We proceed by specifying a model that we can use to test against the null hypothesis. Assume again, for example, that we think a variable X_t may help forecast $\tilde{z}_{t+1}$. Then we can assume the alternative hypothesis

$$\tilde{z}_{t+1} = b_0 + b_1' X_t + \sigma z_{t+1}, \quad \text{with } z_{t+1} \sim \text{i.i.d. } N(0, 1)$$

Then the log-likelihood of a sample of T observations of $\tilde{z}_{t+1}$ under the alternative hypothesis is

$$\ln L\left(b_0, b_1, \sigma^2\right) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\left(\sigma^2\right) - \sum_{t=1}^{T}\left(\frac{\left(\tilde{z}_{t+1} - b_0 - b_1' X_t\right)^2}{2\sigma^2}\right)$$

where we have conditioned on an initial observation.

The parameter estimates $\widehat{b}_0, \widehat{b}_1, \hat{\sigma}^2$ can be obtained from maximum likelihood or, in this simple case, from linear regression. We can then write a likelihood ratio test of correct risk model distribution as

$$LR = -2\left(\ln L(0, 0, 1) - \ln L\left(\widehat{b}_0, \widehat{b}_1, \hat{\sigma}^2\right)\right) \sim \chi^2_{nb+2}$$

where the degrees of freedom in the χ^2 distribution will depend on the number of parameters, nb, in the vector b_1. If we do not have much of an idea about how to choose X_t, then lags of $\tilde{z}_{t+1}$ itself would be obvious choices.

5.1 Backtesting Only the Left Tail of the Distribution

In risk management, we often only really care about forecasting the left tail of the distribution correctly. Testing the entire distribution as we did earlier may lead us to reject risk models that capture the left tail of the distribution well, but not the rest of the distribution. Instead, we should construct a test that directly focuses on assessing the risk model's ability to capture the left tail of the distribution, which contains the largest losses.

Consider restricting attention to the tail of the distribution to the left of the VaR^p_{t+1}—that is, to the $100 \cdot p\%$ largest losses.

If we want to test that the $\tilde{p}_{t+1}$ observations from, for example, the 10% largest losses are themselves uniform, then we can construct a rescaled $\tilde{p}_{t+1}$ variable as

$$\tilde{p}^*_{t+1} = \begin{cases} 10\tilde{p}_{t+1}, \text{ if } \tilde{p}_{t+1} < 0.10 \\ \text{Else not defined} \end{cases}$$

Then we can write the null hypothesis that the risk model provides the correct tail distribution as

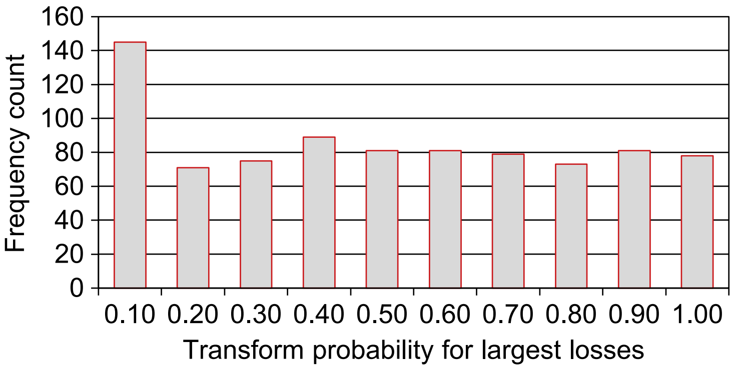
$$H_0 : \tilde{p}^*_{t+1} \sim \text{i.i.d. Uniform}(0, 1)$$

or equivalently

$$H_0 : \tilde{z}^*_{t+1} = \Phi^{-1}\left(\tilde{p}^*_{t+1}\right) \sim \text{i.i.d. } N(0, 1)$$

Figure 13.3 shows the histogram of $\tilde{p}^*_{t+1}$ corresponding to the 10% smallest returns. The data again follow a Student's $t(d)$ distribution with $d = 6$ but the density forecast

Figure 13.3 Histogram of the transform probability from the 10% largest losses.



Notes: We plot the histogram of the transform probability of the 10% largest losses when the returns follow an i.i.d. Student's $t(d)$ distribution with $d = 6$, but they are forecasted by an i.i.d. normal distribution.

model assumes the normal distribution. We have simply zoomed in on the leftmost 10% of the histogram from Figure 13.2. The systematic deviation from a flat histogram is again obvious.

To do formal statistical testing, we can again construct an alternative hypothesis as in

$$\tilde{z}^*_{t+1} = b_0 + b_1' X_t + \sigma z_{t+1}, \quad \text{with } z_{t+1} \sim \text{i.i.d. } N(0,1)$$

for $t+1$ such that $R_{PF,t+1} < -VaR^p_{t+1}$. We can then calculate a likelihood ratio test

$$LR = -2\left(\ln L(0,0,1) - \ln L\left(\hat{b}_0, \hat{b}_1, \hat{\sigma}^2\right)\right) \sim \chi^2_{nb+2}$$

where nb again is the number of elements in the parameter vector b_1.

6 Stress Testing

Due to the practical constraints from managing large portfolios, risk managers often work with relatively short data samples. This can be a serious issue if the historical data available do not adequately reflect the potential risks going forward. The available data may, for example, lack extreme events such as an equity market crash, which occurs very infrequently.

To make up for the inadequacies of the available data, it can be useful to artificially generate extreme scenarios of the main factors driving the portfolio returns (see the exposure mapping discussion in Chapter 7) and then assess the resulting output from the risk model. This is referred to as stress testing, since we are *stressing* the model by exposing it to data different from the data used when specifying and estimating the model.

At first pass, the idea of stress testing may seem vague and ad hoc. Two key issues appear to be (1) how should we interpret the output of the risk model from the stress scenarios, and (2) how should we create the scenarios in the first place? We deal with each of these issues in turn.

6.1 Combining Distributions for Coherent Stress Testing

Standard implementation of stress testing amounts to defining a set of scenarios, running them through the risk model using the current portfolio weights, and if a scenario results in an extreme loss, then the portfolio manager may decide to rebalance the portfolio. Notice how this is very different from deciding to rebalance the portfolio based on an undesirably high *VaR* or Expected Shortfall (*ES*). *VaR* and *ES* are proper probabilistic statements: What is the loss such that I will lose more only 1% of the time (*VaR*)? Or what is the expected loss when I exceed my *VaR* (*ES*)? Standard stress testing does not tell the portfolio manager anything about the probability of the scenario happening, and it is therefore not at all clear what the portfolio rebalancing decision should be. The portfolio manager may end up overreacting to an extreme scenario that occurs with very low probability, and underreact to a less extreme scenario that occurs much more frequently. Unless a probability of occurring is assigned to each scenario, then the portfolio manager really has no idea how to react.

On the other hand, once scenario probabilities are assigned, then stress testing can be very useful. To be explicit, consider a simple example of one stress scenario, which we define as a probability distribution $f_{stress}(\bullet)$ of the vector of factor returns. We simulate a vector of risk factor returns from the risk model, calling it $f(\bullet)$, and we also simulate from the scenario distribution, $f_{stress}(\bullet)$. If we assign a probability α of a draw from the scenario distribution occurring, then we can combine the two distributions as in

$$f_{comb}(\bullet) = \begin{cases} f(\bullet), \text{ with probability } (1-\alpha) \\ f_{stress}(\bullet), \text{ with probability } \alpha \end{cases}$$

Data from the combined distribution is generated by drawing a random variable U_i from a Uniform(0,1) distribution. If U_i is smaller than α, then we draw a return from $f_{stress}(\bullet)$; otherwise we draw it from $f(\bullet)$. The combined distribution can easily be generalized to multiple scenarios, each of which has its own preassigned probability of occurring.

Notice that by simulating from the combined distribution, we are effectively creating a new data set that reflects our available historical data as well our view of the deficiencies of it. The deficiencies are rectified by including data from the stress scenarios in the new combined data set.

Once we have simulated data from the combined data set, we can calculate the *VaR* or *ES* risk measure on the combined data using the previous risk model. If the risk measure is viewed to be inappropriately high then the portfolio can be rebalanced. Notice that now the rebalancing is done taking into account both the magnitude of the stress scenarios and their probability of occurring.

Assigning the probability, α, also allows the risk manager to backtest the *VaR* system using the combined probability distribution $f_{comb}(\bullet)$. Any of these tests can be used to test the risk model using the data drawn from $f_{comb}(\bullet)$. If the risk model, for example, has too many *VaR* violations on the combined data, or if the *VaR* violations come in clusters, then the risk manager should consider respecifying the risk model. Ultimately, the risk manager can use the combined data set to specify and estimate the risk model.

6.2 Choosing Scenarios

Having decided to do stress testing, a key challenge to the risk manager is to create relevant scenarios. The scenarios of interest will typically vary with the type of portfolio under management and with the factor returns applied. The exact choice of scenarios will therefore be situation specific, but in general, certain types of scenarios should be considered. The risk manager ought to do the following:

- *Simulate shocks that are more likely to occur than the historical database suggests*. For example, the available database may contain a few high variance days, but if in general the recent historical period was unusually calm, then the high variance days can simply be replicated in the stress scenario.
- *Simulate shocks that have never occurred but could.* Our available sample may not contain any stock market crashes, but one could occur.
- *Simulate shocks reflecting the possibility that current statistical patterns could break down*. Our available data may contain a relatively low persistence in variance, whereas longer samples suggest that variance is highly persistent. Ignoring the potential persistence in variance could lead to a clustering of large losses going forward.
- *Simulate shocks that reflect structural breaks that could occur*. A prime example in this category would be the sudden float of the previously fixed Thai baht currency in the summer of 1997.

Even if we have identified a set of scenario types, pinpointing the specific scenarios is still difficult. But the long and colorful history of financial crises may serve as a source of inspiration. Examples could include crises set off by political events or natural disasters. For example, the 1995 Nikkei crisis was set off by the Kobe earthquake, and the 1979 oil crisis was rooted in political upheaval. Other crises such as the 1997 Thai baht float and subsequent depreciation mentioned earlier could be the culmination of pressures such as a continuing real appreciation building over time resulting in a loss of international competitiveness.

The effects of market crises can also be very different. They can result in relatively brief market corrections, as was the case after the October 1987 stock market crash, or they can have longer lasting effects, such as the Great Depression in the 1930s. Figure 13.4 depicts the 15 largest daily declines in the Dow Jones Industrial Average during the past 100 years.

Figure 13.4 The fifteen largest one-day percentage declines on the Dow.



Notes: We plot the 15 largest one-day percentage declines in the Dow Jones Industrial Average using data from 1915 through 2010.

Figure 13.4 clearly shows that the October 19, 1987, decline was very large even on a historical scale. We see that the second dip arriving a week later on October 26, 1987 was large by historical standards as well: It was the tenth-largest daily drop. The 2008–2009 financial crisis shows up in Figure 13.4 with three daily drops in the top 15. None of them are in the top 10 however. October–November 1929, which triggered the Great Depression, has four daily drops in the top 10—three of them in the top 5. This bunching in time of historically large daily market drops is quite striking. It strongly suggests that extremely large market drops do not occur randomly but are instead driven by market volatility being extraordinarily high. Carefully modeling volatility dynamics as we did in Chapters 4 and 5 is therefore crucial.

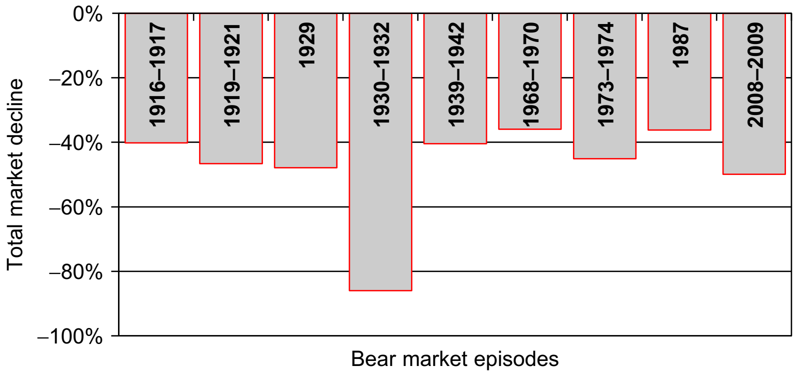
6.3 Stress Testing the Term Structure of Risk

Figure 13.5 shows nine episodes of prolonged market downturn—or bear markets—which we define as at least a 30% decline lasting for at least 50 days. Figure 13.5 shows that the bear market following the 1987 market crash was relatively modest compared to previous episodes. The 2008–2009 bear market during the recent financial crises was relatively large at 50%.

Figure 13.5 suggests that stress testing scenarios should include both rapid corrections, such as the 1987 episode, as well as prolonged downturns that prevailed in 2008–2009.

The Filtered Historical Simulation (or bootstrapping) method developed in Chapter 8 to construct the term structure of risk can be used to stress test the term structure of risk as well. Rather than feeding randomly drawn shocks through the model over time we can feed a path of historical shocks from a stress scenario through the model. The stress scenario can for example be the string of daily shocks observed from September 2008 through March 2009. The outcome of this simulation will show how a stressed market scenario will affect the portfolio under consideration.

Figure 13.5 Bear market episodes in the Dow Jones index.



Notes: We plot the cumulative market decline in nine bear markets defined as cumulative declines of at least 30% lasting at least 50 days. We use daily data from 1915 through 2010 on the Dow Jones Industrial Average.

7 Summary

The backtesting of a risk model can be seen as a final step in model building procedure, and it therefore represents the final chapter in this book. The clustering in time of *VaR* violations as seen in actual commercial bank risk models can pose a serious threat to the financial health of the institution. In this chapter, we therefore developed backtesting procedures capable of capturing such clustering. Backtesting tools were introduced for various risk measures including *VaR*, Expected Shortfall (*ES*), the entire return density, and the left tail of the density.

The more information is provided in the risk measure, the higher statistical power we will have to reject a misspecified risk model. The popular *VaR* risk measure does not, unfortunately, convey a lot of information about the portfolio risk. It tells us a return threshold, which we will only exceed with a certain probability, but it does not tell us about the magnitude of violations that we should expect. The lack of information in the *VaR* makes it harder to backtest. All we can test is that the *VaR* violations fall randomly in time and in the proportion matching the promised coverage rate. Purely from a backtesting perspective, other risk measures such as *ES* and the distribution shape are therefore preferred.

Backtesting ought to be supplemented by stress testing, and we have outlined a framework for doing so. Standard stress testing procedures do not specify the probability with which the scenario under analysis will occur. The failure to specify a probability renders the interpretation of stress testing scenarios very difficult. It is not clear how we should react to a large *VaR* from an extreme scenario unless the likelihood of the scenario occurring is assessed. While it is, of course, difficult to pinpoint the likelihood of extreme events, doing so enables the risk manager to construct a pseudo data set that combines the actual data with the stress scenarios. This combined data set

can be used to backtest the model. Stress testing and backtesting are then done in an integrated fashion.

Further Resources

The *VaR* exceedances from the six U.S. commercial banks in Figure 13.1 are taken from Berkowitz and O'Brien (2002). See also Berkowitz et al. (2011) and O'Brien and Berkowitz (2006). Deng et al. (2008) and Perignon and Smith (2010) present empirical evidence on *VaR*s from an international set of banks.

The *VaR* backtests of unconditional coverage, independence, and conditional coverage are developed in Christoffersen (1998). Kupiec (1995) and Hendricks (1996) restrict attention to unconditional testing. The regression-based approach is used in Christoffersen and Diebold (2000). Christoffersen and Pelletier (2004) and Candelon et al. (2011) construct tests based on the duration of time between *VaR* hits. Campbell (2007) surveys the available backtesting procedures.

Christoffersen and Pelletier (2004) discuss the details in implementing the Monte Carlo simulated P-values, which were originally derived by Dufour (2006).

Christoffersen et al. (2001), Giacomini and Komunjer (2005), and Perignon and Smith (2008) develop tests for comparing different *VaR* models. Andreou and Ghysels (2006) consider ways of detecting structural breaks in the return process for the purpose of financial risk management. For a regulatory perspective on backtesting, see Lopez (1999) and Kerkhof and Melenberg (2004). Lopez and Saidenberg (2000) focus on credit risk models. Zumbach (2006) considers different horizons.

Engle and Manganelli (2004), Escanciano and Olmo (2010, 2011), and Gaglianone et al. (2011) suggest quantile-regression approaches and allow for parameter estimation error.

Procedures for backtesting the Expected Shortfall risk measures can be found in McNeil and Frey (2000) and Angelidis and Degiannakis (2007).

Graphical tools for assessing the quality of density forecasts are suggested in Diebold et al. (1998). Crnkovic and Drachman (1996), Berkowitz (2001), and Bontemps and Meddahi (2005) establish formal statistical density evaluation tests, and Berkowitz (2001), in addition, suggested focusing attention to backtesting the left tail of the density. See also the survey in Tay and Wallis (2007) and Corradi and Swanson (2006).

The coherent framework for stress testing is spelled out in Berkowitz (2000). See also Kupiec (1998), Longin (2000), and Alexander and Sheedy (2008). Rebonato (2010) takes a Bayesian approach and devotes an entire book to the topic of stress testing.

The May 1998 issue of the *World Economic Outlook*, published by the International Monetary Fund (see www.imf.org), contains a useful discussion of financial crises during the past quarter of a century. Kindleberger and Aliber (2000) take an even longer historical view.

References

Alexander, C., Sheedy, E., 2008. Developing a stress testing framework based on market risk models. J. Bank. Finance 32, 2220–2236.

Andreou, E., Ghysels, E., 2006. Monitoring distortions in financial markets. J. Econom. 135, 77–124.

Angelidis, T., Degiannakis, S.A., 2007. Backtesting VaR models: An expected shortfall approach. http://ssrn.com/paper=898473.

Berkowitz, J., 2000. A coherent framework for stress testing. J. Risk Winter 2, 1–11.

Berkowitz, J., 2001. Testing density forecasts, applications to risk management. J. Bus. Econ. Stat. 19, 465–474.

Berkowitz, J., Christoffersen, P., Pelletier, D., 2011. Evaluating value-at-risk models with desk-level data. Manag. Sci. forthcoming.

Berkowitz, J., O'Brien, J., 2002. How accurate are the value-at-risk models at commercial banks? J. Finance 57, 1093–1112.

Bontemps, C., Meddahi, N., 2005. Testing normality: A GMM approach. J. Econom. 124, 149–186.

Campbell, S., 2007. A review of backtesting and backtesting procedures. J. Risk 9, Winter, 1–17.

Candelon, B., Colletaz, G., Hurlin, C., Tokpavi, S., 2011. Backtesting value-at-risk: A GMM duration-based test. J. Financ. Econom. 9, 314–343.

Christoffersen, P., 1998. Evaluating interval forecasts. Int. Econ. Rev. 39, 841–862.

Christoffersen, P., Diebold, F., 2000. How relevant is volatility forecasting for financial risk management? Rev. Econ. Stat. 82, 12–22.

Christoffersen, P., Hahn, J., Inoue, A., 2001. Testing and comparing value-at-risk measures. J. Empir. Finance 8, 325–342.

Christoffersen, P., Pelletier, D., 2004. Backtesting portfolio risk measures: A duration-based approach. J. Financ. Econom. 2, 84–108.

Corradi, V., Swanson, N., 2006. Predictive density evaluation. In: Elliot, G., Gringer, C., Timmcrmann, A. (Eds.), Handbook of Economic Forecasting. Elsevier, North Holland, vol. 1. pp. 197–284.

Crnkovic, C., Drachman, J., 1996. Quality control. Risk September 9, 138–143.

Deng, Z, Perignon, C., Wang, Z., 2008. Do banks overstate their value-at-risk? J. Bank. Finance 32, 783–794.

Diebold, F.X., Gunther, T., Tay, A., 1998. Evaluating density forecasts, with applications to financial risk management. Int. Econ. Rev. 39, 863–883.

Dufour, J.-M., 2006. Monte Carlo tests with nuisance parameters: A general approach to finite sample inference and non-standard asymptotics. J. Econom. 133, 443–477.

Engle, R., Manganelli, S., 2004. CAViaR: Conditional value at risk by quantile regression. J. Bus. Econ. Stat. 22, 367–381.

Escanciano, J., Olmo, J., 2010. Backtesting parametric value-at-risk with estimation risk. J. Bus. Econ. Stat. 28, 36–51.

Escanciano, J., Olmo, J., 2011. Robust backtesting tests for value-at-risk models. J. Financ. Econom. 9, 132–161.

Gaglianone, W., Lima, L., Linton, O., 2011. Evaluating value-at-risk models via quantile regression. J. Bus. Econ. Stat. 29, 150–160.

Giacomini, R., Komunjer, I., 2005. Evaluation and combination of conditional quantile forecasts. J. Bus. Econ. Stat. 23, 416–431.

Hendricks, D., 1996. Evaluation of value-at-risk models using historical data. Econ. Policy Rev., Federal Reserve Bank of New York 2, 39–69.
International Monetary Fund, 1998. World Economic Outlook, May. IMF, Washington, DC. Available from: www.IMF.org.
Kerkhof, J., Melenberg, B., 2004. Backtesting for risk-based regulatory capital. J. Bank. Finance 28, 1845–1865.
Kindleberger, C., Aliber, R., 2000. Manias, Panics and Crashes: A History of Financial Crisis. John Wiley and Sons, New York.
Kupiec, P., 1995. Techniques for verifying the accuracy of risk measurement models. J. Derivatives 3, 73–84.
Kupiec, P., 1998. Stress testing in a value at risk framework. J. Derivatives 6, 7–24.
Longin, F.M., 2000. From value at risk to stress testing: The extreme value approach. J. Bank. Finance 24, 1097–1130.
Lopez, J., 1999. Regulatory evaluation of value-at-risk models. J. Risk 1, 37–64.
Lopez, J., Saidenberg, M., 2000. Evaluating credit risk models. J. Bank. Finance 24, 151–165.
McNeil, A., Frey, R., 2000. Estimation of tail-related risk measures for heteroskedastic financial time series: An extreme value approach. J. Empir. Finance 7, 271–300.
O'Brien, J., Berkowitz, J., 2006. Bank trading revenues, VaR and market risk. In: Stulz, R., Carey, M. (Eds.), The Risks of Financial Institutions. University of Chicago Press for NBER, Chicago, Illinois, pp. 59–102.
Perignon, C., Smith, D., 2008. A new approach to comparing VaR estimation methods. J. Derivatives 15, 54–66.
Perignon, C., Smith, D., 2010. The level and quality of value-at-risk disclosure by commercial banks. J. Bank. Finance 34, 362–377.
Rebonato, R., 2010. Coherent Stress Testing: A Bayesian Approach to the Analysis of Financial Stress. John Wiley and Sons, Chichester, West Sussex, UK.
Tay, A., Wallis, K., 2007. Density forecasting: A survey. In: Clements, M., Hendry, D. (Eds.), A Companion to Economic Forecasting, Blackwell Publishing Malden, MA, pp. 45–68.
Zumbach, G., 2006. Backtesting risk methodologies from one day to one year. J. Risk 11, 55–91.

Empirical Exercises

Open the Chapter13Data.xlsx file from the web site.

1. Compute the daily variance of the returns on the S&P 500 using the RiskMetrics approach.
2. Compute the 1% and 5% 1-day Value-at-Risk for each day using RiskMetrics and Historical Simulation with 500 observations.
3. For the 1% and 5% value at risk, calculate the indicator "hit" sequence for both RiskMetrics and Historical Simulation models. The hit sequence takes on the value 1 if the return is below the (negative of the) *VaR* and 0 otherwise.
4. Calculate the LR_{uc}, LR_{ind}, and LR_{cc} tests on the hit sequence from the RiskMetrics and Historical Simulation models. (*Excel hint*: Use the CHIINV function.) Can you reject the *VaR* model using a 10% significance level?
5. Using the RiskMetrics variances calculated in exercise 1, compute the uniform transform variable. Plot the histogram of the uniform variable. Does it look flat?
6. Transform the uniform variable to a normal variable using the inverse cumulative density function (CDF) of the normal distribution. Plot the histogram of the normal variable. What

is the mean, standard deviation, skewness, and kurtosis? Does the variable appear to be normally distributed?

7. Take all the values of the uniform variable that are less than or equal to 0.1. Multiply each number by 10. Plot the histogram of this new uniform variable. Does it look flat? Why should it?

8. Transform the new uniform variable to a normal variable using the inverse CDF of the normal distribution. Plot the histogram of the normal variable. What is the mean, standard deviation, skewness, and kurtosis? Does the variable appear to be normally distributed?

The answers to these exercises can be found in the Chapter13Results.xlsx file on the companion site.

For more information see the companion site at
http://www.elsevierdirect.com/companions/9780123744487

Index

Note: Page numbers followed by "*f*" indicates figures